Gossip and Distributed Kalman Filtering: Weak Consensus under Weak Detectability

Soummya Kar† and José M. F. Moura†

Abstract

The paper presents the gossip interactive Kalman filter (GIKF) for distributed Kalman filtering for networked systems and sensor networks, where inter-sensor communication and observations occur at the same time-scale. The communication among sensors is random; each sensor occasionally exchanges its filtering state information with a neighbor depending on the availability of the appropriate network link. We show that under a weak distributed detectability condition: 1) the GIKF error process remains stochastically bounded, irrespective of the instability properties of the random process dynamics; and 2) the network achieves *weak consensus*, i.e., the conditional estimation error covariance at a (uniformly) randomly selected sensor converges in distribution to a unique invariant measure on the space of positive semi-definite matrices (independent of the initial state.) To prove these results, we interpret the filtered states (estimates and error covariances) at each node in the GIKF as stochastic particles with local interactions. We analyze the asymptotic properties of the error process by studying as a random dynamical system the associated switched (random) Riccati equation, the switching being dictated by a non-stationary Markov chain on the network graph.

Keywords: Gossip, Kalman filter, consensus, random dynamical systems, random algebraic Riccati equation.

I. Introduction

A. Background and Motivation

This paper presents the Gossip Interactive Kalman Filtering (GIKF). GIKF is a linear distributed estimator that filters noisy observations of a random process measured by a sparsely connected sensor network. Each sensor observes only a portion of the process, such that, acting alone, no sensor can resolve the signal. GIKF is fundamentally different from other distributed implementations of the Kalman filter ([1], [2], [3], [4]) that employ some form of linear consensus on the sensor observations or estimates; in contrast, GIKF involves communication and observation sampling at the same time scale. GIKF runs at each sensor a local copy of the Kalman filter and achieves collaboration through occasional asynchronous state swaps between sensors at random time instants. At the random times when a sensor communicates with one of its randomly selected neighbors, the sensor swaps its previous state (its local Kalman filter state estimate and conditional error covariance) with the state of its neighbor, before processing the current observation. In other words, when communication is established, a sensor updates the state it receives from its neighbor with its present observation; otherwise, it updates its own previous state. Such collaboration or information exchange through state swapping is asynchronous over the network and occurs occasionally, dictated by the random network topology. Indeed, due to inherent environmental randomness, the underlying medium access control (MAC) protocol is randomized and often not known at the local sensor level. We assume that the sensor network uses a generic random communication protocol, see Section II, that subsumes the widely used gossiping protocol for real time embedded architectures, [5], and the graph matching based communication protocols for internet architectures, [6].

The paper establishes GIKF and studies its error properties. We define a weak distributed detectability condition[1] under which we show: 1) the GIKF error process remains stochastically bounded, irrespective of the instability properties of the random process dynamics; and 2) the network achieves weak consensus, i.e., the conditional estimation error covariance at a (uniformly) randomly selected sensor converges in distribution to a unique invariant measure on the space of positive semi-definite matrices (independent of the initial state.) To prove these results, we interpret the filtered states at each node in the GIKF as stochastic particles with local interactions and analyze the asymptotic properties of the error process by studying as a random dynamical system the switched (random) Riccati equation, the switching being dictated by a non-stationary Markov chain on the network graph.

To study the information flows in the GIKF, we interpret the filtering states at each node as stochastic particles with controlled interactions. To prove the stochastic boundedness of the error process and the network weak consensus, we focus on these traveling states, which we refer to as tokens or particles, and not on the sequence of conditional error covariances at each sensor, which is not Markov. This particle point of view is reminiscent of the approach taken in fluid dynamics of studying the transport of

---

[1]This condition is required even by a centralized estimator (having access to all sensor observations over all time) to yield an estimate with bounded error (for unstable systems.)

a particle as it travels in the fluid (Lagrangian coordinates) rather than studying the transport at a fixed coordinate in space (Eulerian coordinates), [7]. We show that the sequence of traveling states or particles evolves according to a switched system of random Riccati operators, the switching being dictated by a non-stationary Markov chain on the graph. A key contribution is the analysis of the resulting random Riccati equation (RRE). In this context, we note that the RRE arises in the literature in several practical filtering and control formulations with non-classical information. Prior work ([8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]) mostly address qualitative properties of the RRE in terms of moment stability, whereas recent approaches focus on understanding the limit behavior in terms of weak convergence ([19], [20], [21], [22], see also [23]). In this paper, we utilize a random dynamical systems formulation of the RRE; however, in contrast with our work in [19], [20], the switching sequence is no longer stationary. Several approximation arguments of independent interest are developed to tackle this non-stationary behavior and to establish the asymptotic distributional properties of the RRE.

To summarize, the paper addresses two fundamental concerns in collaborative estimation in random environments. It introduces *distributed observability* for linear dynamical estimation and addresses the question of *minimal* observation pattern (i.e., what should be the minimal number of sensors and what should they observe,) so that there exists a *successful* filtering scheme. The weak detectability condition (introduced in Section II-A) resolves this question through the existence of a full rank network Grammian. We show that satisfaction of the weak detectability condition leads to stochastic boundedness of the conditional filtering error at each sensor, irrespective of observability of individual sensors. The second concern addressed in the paper is that of robust information flow, which seeks to address the minimal communication required to maintain consistent (asymptotically) information dissemination in the network. The weak connectedness assumption formulated in Section II-A quantifies the rate of information flow (in random communication environments) as the mixing time of a particle undergoing a random walk in the network with appropriate statistics. The positive recurrence of this Markov chain translates to information dissemination at a sufficient rate to cope with the (possible) instability in signal dynamics and leads to weak consensus of the filtering errors. The notion of weak consensus introduced in the paper is the best form of *consensus* possible in such a setup because, as opposed to familiar scenarios (average computation/static parameter estimation,) in a dynamic situation it is not possible to accomplish almost sure (pathwise) consensus of the estimate or error processes. On the contrary, the weak consensus we establish shows that the error processes at different sensors converge in distribution to the same invariant measure. We do not characterize here this invariant measure as a function of the communication and observation policies; instead, we resolve the minimal conditions for the existence of such an invariant measure and hence conditions for the stability of the filtering error processes.

We briefly summarize the organization of the rest of the paper. Subsection I-B sets up notation and background material to be used in the paper. Section II sets-up the problem and introduces the GIKF algorithm together with the observability and connectivity assumptions in Subsection II-A. An interactive particle interpretation and important preliminary results are in Subsection II-B. The main results regarding

the asymptotic properties of the GIKF are stated (without proof) and interpreted in Section III. To prove these results, we provide first in Section IV a random dynamic system (RDS) formulation of the switching iterates of the random Riccati equation arising in the GIKF. Appendix A recalls facts and results on random dynamical systems (RDS) needed in this Section. The main results of the paper are proved in Section VI. Two technical Lemmas are proven in Appendix B. Finally, Section VII concludes the paper.

B. Notation and Preliminaries

Let $\mathbb{R}$ be the reals; $\mathbb{R}^M$, the M-dimensional Euclidean space; $\mathbb{T}$, the integers; $\mathbb{T}_+$, the non-negative integers; $\mathbb{N}$, the natural numbers; and $\mathcal{X}$, a generic space. For $B \subset \mathcal{X}$, $\mathbb{I}_B : \mathcal{X} \longmapsto \{0, 1\}$ is the indicator function, i.e., 1 when its argument is in B and zero otherwise; and $\text{id}_\mathcal{X}$ is the identity function on $\mathcal{X}$.

Cones in partially ordered Banach spaces. We summarize facts and definitions on the structure of cones in partially ordered Banach spaces. Let V be a Banach space (over the field of the reals) with a closed (w.r.t. the Banach space norm) convex cone V_+ and assume $V_+ \cap (-V_+) = \{0\}$. The cone V_+ induces a partial order in V, namely, for $X, Y \in V$, we write $X \preceq Y$, if $Y - X \in V_+$. In case $X \preceq Y$ and $X \neq Y$, we write $X \prec Y$. The cone V_+ is called solid, if it has a non-empty interior $\text{int}\, V_+$; in that case, V_+ defines a strong ordering in V, and we write $X \ll Y$, if $Y - X \in \text{int}\, V_+$. The cone V_+ is normal if the norm $\| \cdot \|$ of V is semi-monotone, i.e., $\exists\, c > 0$, s.t. $0 \preceq X \preceq Y \Rightarrow \|X\| \leq c\|Y\|$. There are various equivalent characterizations of normality, of which we note that the normality of V_+ ensures that the topology in V induced by the Banach space norm is compatible with the ordering induced by V_+, in the sense that any norm-bounded set $B \subset V$ is contained in a conic interval of the form $[X, Y]$, where $X, Y \in V$. Finally, a cone is said to be minihedral, if every order-bounded (both upper and lower bounded) finite set $B \subset V$ has a supremum (here bounds are w.r.t. the partial order.)

We focus on the separable Banach space of symmetric $n \times n$ matrices, $\mathbb{S}^n$, equipped with the induced 2-norm. The subset $\mathbb{S}_+^N$ of positive semidefinite matrices is a closed, convex, solid, normal, minihedral cone in $\mathbb{S}^n$, with non-empty interior $\mathbb{S}_{++}^N$, the set of positive definite matrices. The conventions above denote the partial and strong ordering in $\mathbb{S}^n$ induced by $\mathbb{S}_+^N$.

Probability measures on metric spaces: Let: $(\mathcal{X}, d_\mathcal{X})$ a complete separable metric space $\mathcal{X}$ with metric $d_\mathcal{X}$; $\mathbb{B}(\mathcal{X})$ its Borel algebra; $B(\mathcal{X})$ the Banach space of real-valued bounded functions on $\mathcal{X}$, equipped with the sup-norm, i.e., $f \in B(\mathcal{X})$, $\|f\| = \sup_{x \in \mathcal{X}} |f(x)|$; and $C_b(\mathcal{X})$ the subspace of $B(\mathcal{X})$ of continuous functions. For $x \in \mathcal{X}$, the open ball of radius $\varepsilon > 0$ centered at x is denoted by $B_\varepsilon(x)$, i.e., $B_\varepsilon(x) = \{y \in \mathcal{X} \mid d_\mathcal{X}(y, x) < \varepsilon\}$. For any set $\Gamma \subset \mathcal{X}$, the open ε-neighborhood of Γ is given by $\Gamma_\varepsilon = \{y \in \mathcal{X} \mid \inf_{x \in \Gamma} d_\mathcal{X}(y, x) < \varepsilon\}$. It can be shown that Γ_ε is an open set.

Let $\mathcal{P}(\mathcal{X})$ be the set of probability measures on $\mathcal{X}$. A sequence $\{\mu_t\}_{t \in \mathbb{T}_+}$ of probability measures in $\mathcal{P}(\mathcal{X})$ converges weakly to $\mu \in \mathcal{P}(\mathcal{X})$ if $\lim_{t \to \infty} < f, \mu_t > = < f, \mu >, \ \forall\, f \in C_b(\mathcal{X})$. By Portmanteau's theorem, the above is equivalent to any one of the following:

[i] For all closed $F \in \mathbb{B}(\mathcal{X})$ $\qquad \limsup_{t \to \infty} \mu_t(F) \leq \mu(F)$

[ii] For all open $O \in \mathbb{B}(\mathcal{X})$ $\qquad \liminf_{t \to \infty} \mu_t(O) \geq \mu(O)$

Weak convergence is denoted by $\mu_t \Longrightarrow \mu$ and is also referred to as convergence in distribution. The weak topology on $\mathcal{P}(\mathcal{X})$ generated by weak convergence can be metrized. In particular, e.g., [24], one has the Prohorov metric d_p on $\mathcal{P}(\mathcal{X})$, such that the metric space $(\mathcal{P}(\mathcal{X}), d_p)$ is complete, separable, and a sequence $\{\mu_t\}_{t \in \mathbb{T}_+}$ in $\mathcal{P}(\mathcal{X})$ converges weakly to μ in $\mathcal{P}(\mathcal{X})$ *iff* $\lim_{t \to \infty} d_p(\mu_t, \mu) = 0$. The distance between two probability measures μ_1, μ_2 in $\mathcal{P}(\mathcal{X})$ is computed as:

$$d_P(\mu_1, \mu_2) = \inf \{\varepsilon > 0 \mid \mu_1(\mathcal{F}) \leq \mu_2(\mathcal{F}_\varepsilon) + \varepsilon, \quad \forall \text{ closed set } \mathcal{F}\} \tag{1}$$

II. GOSSIP INTERACTIVE KALMAN FILTER (GIKF)

A. Problem setup

Signal/Observation Model We consider a discrete-time linear dynamical system observed by a network of N sensors. The signal model is:

$$\mathbf{x}_{t+1} = \mathcal{F}\mathbf{x}_t + \mathbf{w}_t \tag{2}$$

where $\mathbf{x}_t \in \mathbb{R}^M$ is the signal (state) vector with initial state $\mathbf{x}_0$ distributed as a zero mean Gaussian vector with covariance $\widehat{P}_0$ and the system noise $\{\mathbf{w}_t\}$ is an uncorrelated zero mean Gaussian sequence independent of $\mathbf{x}_0$ with covariance $\mathcal{Q}$. The observation at the n-th sensor $\mathbf{y}_t^n \in \mathbb{R}^{m_n}$ at time t is:

$$\mathbf{y}_t^n = \mathcal{C}_n \mathbf{x}_t + \mathbf{v}_t^n \tag{3}$$

where $\mathcal{C}_n \in \mathbb{R}^{m_n \times M}$ and $\{\mathbf{v}_t^n\}$ is an uncorrelated zero mean Gaussian observation noise sequence with covariance $\mathcal{R}_n \gg \mathbf{0}$. Also, the noise sequences at different sensors are independent of each other, the system noise process and the initial system state. Because of the limited capability of the sensors, typically the dimension of $\mathbf{y}_t^n$ is much smaller than that of the signal process and the observation process at each sensor is not sufficient to make the pair $\{\mathbf{x}_t, \mathbf{y}_t^n\}$ observable[2]. We envision a totally distributed application where a reliable estimate of the signal process is required at each sensor.[3] The sensors achieve collaboration with each other by means of occasional communication with their *neighbors*, whereby they exchange their filtering states (to be defined precisely.) We assume that time is slotted and inter-sensor communication and sensing (observation) take place at the same time-scale.

Communication Model Communication among sensors is constrained by several factors including proximity, transmit power, and receiving capabilities. We model the underlying communication structure of the network in terms of an undirected graph $(V, \mathcal{E})$ where V denotes the set of N sensors and $\mathcal{E}$ is the set of edges or allowable communication links between the sensors. The notation $n \sim l$ indicates

---

[2]It is possible that some of the sensors have no observation capabilities, i.e., the corresponding C_n is a zero matrix. Thus the formulation easily carries over to networks of heterogeneous agents, consisting of 'sensors' which actually sense the field of interest and actuators, which implement local control actions based on the estimated field.

[3]The term sensor network here refers to a network of *agents* (possibly distributed over a geographical region) with varied functionalities. For example, some agents may be physical sensors while others may be remote actuators, in which case, the corresponding observation matrix C_n is identically zero. In this paper, we use the term sensor to denote a generic network agent.

that sensors n and l can communicate, i.e., $\mathcal{E}$ contains the undirected edge (n, l). The graph can be represented in terms of its $N \times N$ symmetric adjacency matrix $\mathcal{A}$:

$$\mathcal{A}_{nl} = \begin{cases} 1 & \text{if } (n, l) \in \mathcal{E} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

We assume that the diagonal elements of $\mathcal{A}$ are identically 1, indicating that a sensor n can always communicate to itself. Note, that $\mathcal{E}$ is the maximal allowable set of links in the network at any time, however, at a particular instant, each sensor may choose to communicate only to a fraction of its neighbors. The exact communication protocol is not so important for the analysis, as long as some *weak* connectivity assumptions are satisfied. For definiteness, we assume the following generic communication model, which subsumes the widely used gossiping protocol for real time embedded architectures ([5]) and the graph matching based communication protocols for internet architectures ([6].) We make this precise in the following, which we generalize later. Define the set $\mathcal{M}$ of symmetric 0-1 $N \times N$ matrices:

$$\mathcal{M} = \left\{ A \mid \mathbf{1}^T A = \mathbf{1}^T, \ \ A\mathbf{1} = \mathbf{1}, \ \ A \leq \mathcal{E} \right\} \tag{5}$$

In other words, $\mathcal{M}$ is the set of adjacency matrices, such that, every node is incident to exactly one edge (including self edges) and allowable edges are only those included in E.[4] Let $\mathcal{D}$ be a probability distribution on the space $\mathcal{M}$. The sequence of time-varying adjacency matrices, $\{A(t)\}_{t \in \mathbb{N}}$, governing the inter-sensor communication, is then an i.i.d. sequence in $\mathcal{M}$ with distribution $\mathcal{D}$ and independent of the signal and observation processes.[5] We make the following assumption of connectivity on the average:

Assumption C.1: Define the symmetric stochastic matrix $\overline{A}$ as

$$\overline{A} = \mathbb{E}\left[A(t)\right] = \int_{\mathcal{M}} A d\mathcal{D}(A) \tag{6}$$

The matrix $\overline{A}$ is assumed to be irreducible and aperiodic.

Remark 1 The stochasticity of $\overline{A}$ is inherited from that of the elements of $\mathcal{M}$. We are not concerned with the properties of the distribution $\mathcal{D}$ as long as the *weak* connectivity assumption above is satisfied. The issue of $\overline{A}$ being irreducible depends both on the set of allowable edges $\mathcal{E}$ and the distribution $\mathcal{D}$. We do not pursue that question in detail here. However, to show the applicability of Assumption **C.1** and justify the notion of weak connectivity, we note that such a distribution $\mathcal{D}$ always exists if the graph $(V, \mathcal{E})$ is connected. We give a Markov chain interpretation of the mean adjacency matrix $\overline{A}$, which will be helpful for the analysis to follow. The matrix $\overline{A}$ can be associated to the transition kernel of a time-homogeneous Markov chain on the state space V. Since the state space V is finite, the irreducibility of $\overline{A}$ suggests that the resulting Markov chain is positive recurrent. Due to symmetricity, the Markov chain is reversible with unique invariant distribution π on V, where π is the discrete uniform distribution on V.

---

[4]The set $\mathcal{M}$ is always non-empty, in particular the $N \times N$ identity matrix $I_N \in \mathcal{M}$.

[5]For convenience of presentation, we assume that $A(0) = I_N$, although communication starts at slot $t = 1$.

Observability Conditions: Weak Detectability Successful filtering even in the centralized setting (assuming all the sensors can forward their observations at all time to a fusion center) requires some form of detectability and stabilizability. In the present distributed setting we impose the following weak assumptions on the signal/observation model:

Stabilizability: Assumption S.1 The pair $(\mathcal{F}, \mathcal{Q}^{1/2})$ is stabilizable. The non-degeneracy of $\mathcal{Q}$ ensures this.

For distributed detectability, we assume the following:

Weak Detectability: Assumption D.1 There exists a walk[6] of length $\ell \geq 1$, $(n_1, n_2, \cdots, n_\ell)$, covering the N nodes, such that, the matrix $\sum_{i=1}^{\ell} \left(\mathcal{F}^{i-1}\right)^T \mathcal{C}_{n_i}^T \mathcal{C}_{n_i} \mathcal{F}^{i-1}$ is invertible.

Remark 2 Note, as permitted by the general definition of a walk, the sequence $(n_1, n_2, \cdots, n_\ell)$ may consist of repeated vertices and, in particular, self-loops (if permitted by $\overline{A}$.)

Remark 3 When $\mathcal{F}$ is invertible, **D.1** may be replaced by the full rank of

$$\mathcal{G} = \sum_{n=1}^{N} \mathcal{C}_n^T \mathcal{C}_n \tag{7}$$

Indeed, by the irreducibility of $\overline{A}$ (equivalently, by the connectivity of the graph induced by $\overline{A}$,) we can find a walk $(n_1, n_2, \cdots, n_\ell)$ of length $\ell \geq N$, which covers the network, i.e., visits each node at least once. Hence, if $\mathcal{F}$ is invertible and (7) holds, it follows that the matrix $\sum_{i=1}^{\ell} \left(\mathcal{F}^{i-1}\right)^T \mathcal{C}_{n_i}^T \mathcal{C}_{n_i} \mathcal{F}^{i-1}$ corresponding to this walk is invertible leading to Assumption **D.1**.

Remark 4 From the positive definiteness of the measurement noise matrices $\mathcal{R}_n$, it follows that under **D.1**, the matrix $\sum_{i=1}^{\ell} \left(\mathcal{F}^{i-1}\right)^T \mathcal{C}_{n_i}^T \mathcal{R}_{n_i}^{-1} \mathcal{C}_{n_i} \mathcal{F}^{i-1}$ is invertible.

Remark 5 Assumption **D.1** is minimal, in the sense, that, even in a centralized setting (a center has access to all the sensor observations over all time,) it is required to ensure detectability for arbitrary choice of the matrix F governing the signal dynamics. This justifies the nomenclature weak detectability.

Algorithm GIKF We now present the algorithm GIKF (gossip based interacting Kalman filter) for distributed estimation of the signal process $\mathbf{x}_t$ over time. We start by introducing notation. Let the filter at sensor n be initialized with the pair $\left(\widehat{\mathbf{x}}_{0|-1}, \widehat{P}_0\right)$, where $\widehat{\mathbf{x}}_{0|-1}$ denotes the prior estimate of $\mathbf{x}_0$ (with no observation information) and $\widehat{P}_0$ the corresponding error covariance. Also, by $(\widehat{\mathbf{x}}_{t|t-1}^n, \widehat{P}_t^n)$ denote the estimate at sensor n of $\mathbf{x}_t$ based on information[7] till time $t-1$ and the corresponding conditional error covariance, respectively. The pair $\left(\widehat{\mathbf{x}}_{t|t-1}^n, \widehat{P}_t^n\right)$ is also referred to as the state of sensor n at time $t-1$. To define the estimate update rule for the GIKF, let $\rightarrow (n, t)$ be the neighbor of sensor n at time t

---

[6]A walk in this context is defined w.r.t. the graph induced by the non-zero entries of the matrix $\overline{A}$.

[7]The information at sensor n till (and including) time t corresponds to the sequence of observations $\{\mathbf{y}_s^n\}_{0 \leq s \leq t}$ obtained at the sensor and the information received by data exchange between its neighboring senors.

w.r.t. the adjacency matrix [8] $A(t)$. We assume that all inter-sensor communication for time t occurs at the beginning of the slot, whereby communicating sensors swap their previous states, i.e., if at time t, $\to (n,t) = l$, sensor n replaces its previous state $\left(\widehat{\mathbf{x}}_{t|t-1}^n, \widehat{P}_t^n\right)$ by $\left(\widehat{\mathbf{x}}_{t|t-1}^l, \widehat{P}_t^l\right)$ and sensor l replaces its previous state $\left(\widehat{\mathbf{x}}_{t|t-1}^l, \widehat{P}_t^l\right)$ by $\left(\widehat{\mathbf{x}}_{t|t-1}^n, \widehat{P}_t^n\right)$. The estimate update at sensor n at the end of the slot (after the communication and observation tasks have been completed) is:

$$\widehat{\mathbf{x}}_{t+1|t}^n = \mathbb{E}\left[\mathbf{x}_{t+1} \;\middle|\; \widehat{\mathbf{x}}_{t|t-1}^{\to(n,t)}, \widehat{P}_t^{\to(n,t)}, \mathbf{y}_t^n\right] \tag{8}$$

$$\widehat{P}_{t+1}^n = \mathbb{E}\left[\left(\mathbf{x}_{t+1} - \widehat{\mathbf{x}}_{t+1|t}^n\right)\left(\mathbf{x}_{t+1} - \widehat{\mathbf{x}}_{t+1|t}^n\right)^T \;\middle|\; \widehat{\mathbf{x}}_{t|t-1}^{\to(n,t)}, \widehat{P}_t^{\to(n,t)}, \mathbf{y}_t^n\right] \tag{9}$$

Due to conditional Gaussianity, the filtering steps above can be implemented through the time-varying Kalman filter recursions, and it follows that the sequence $\left\{\widehat{P}_t^n\right\}$ of conditional predicted error covariance matrices at sensor n satisfies the Riccati recursion:

$$\widehat{P}_{t+1}^n = \mathcal{F}\widehat{P}_t^{\to(n,t)}\mathcal{F}^T + \mathcal{Q} - \mathcal{F}\widehat{P}_t^{\to(n,t)}\mathcal{C}_n^T \left(\mathcal{C}_n\widehat{P}_t^{\to(n,t)}\mathcal{C}_n^T + \mathcal{R}_n\right)^{-1} \mathcal{C}_n\widehat{P}_t^{\to(n,t)}\mathcal{F}^T \tag{10}$$

Note that the sequence $\left\{\widehat{P}_t^n\right\}$ is random, due to the random neighborhood selection function $\to (n,t)$. The goal of the paper is to study asymptotic properties of the sequence of random conditional error covariance matrices $\left\{\widehat{P}_t^n\right\}$ at every sensor n and show in what sense they reach *consensus*, so that, in the limit of large time, every sensor provides an equally good (stable in the sense of estimation error) estimate of the signal process.

B. An Interacting Particle Representation

To compactify the notation in eqn. (10), we define the functions $f_n : \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ for $n = 1, \cdots, N$ defining the respective Riccati operators[9]:

$$f_n(X) = \mathcal{F}X\mathcal{F}^T + \mathcal{Q} - \mathcal{F}X\mathcal{C}_n^T \left(\mathcal{C}_n X\mathcal{C}_n^T + \mathcal{R}_n\right)^{-1} \mathcal{C}_n X\mathcal{F}^T \tag{11}$$

Recall the sequence $\{\to (n,t)\}_{t\in\mathbb{T}_+}$ of neighborhoods of sensor n. The sequence of conditional error covariance matrices $\{P_t^n\}_{t\in\mathbb{T}_+}$ at sensor n then evolves according to

$$\widehat{P}_{t+1}^n = f_n\left(\widehat{P}_t^{\to(n,t)}\right) \tag{12}$$

The sequence $\left\{\widehat{P}_t^n\right\}$ is non-Markov (not even semi-Markov given the random adjacency matrix sequence $\{A(t)\}$,) as $\widehat{P}_{t+1}^n$ at time t is a random functional of the conditional error covariance at time $t-1$ of the sensor $\to (n,t)$, which, in general, is different from sensor n. This makes the evolution of the sequence $\left\{\widehat{P}_t^n\right\}$ difficult to track. To overcome this, we give the following interacting particle interpretation of

---

[8]Note that $n(t)$ is unambiguously defined as $A(t)$ is a matching matrix, and also by symmetry we have $\to (\to (n,t),t) = n$. It is possible that $\to (n,t) = n$, in which case the graph corresponding to $A(t)$ has a self-loop at node n.

[9]In case a sensor does not observe, i.e., $C_n = 0$, then the corresponding Riccati operator f_n in eqn. (11) reduces to the Lyapunov operator.

the conditional error covariance evolution, which naturally leads us to track semi-Markov sequences of conditional error covariance matrices from which we can completely characterize the evolution of the desired covariance sequences $\left\{ \widehat{P}_t^n \right\}$ for $n = 1, \cdots, N$.

To this end, we note that the link formation process given by the sequence $\{A(t)\}$ can be represented in terms of N particles moving on the graph as identical Markov chains. The state of the n-th particle is denoted by $p_n(t)$, and the sequence $\{p_n(t)\}_{t \in \mathbb{T}_+}$ takes values in $[1, \cdots, N]$. The evolution of the n-th particle is given as follows:

$$p_n(t) = \to (p_n(t-1), t), \quad p_n(0) = n \tag{13}$$

Recall the (random) neighborhood selection $\to (n, t)$. Thus, the n-th particle can be viewed as originating from node n at time 0 and then traveling on the graph (possibly changing its location at each time) according to the link formation process $\{A(t)\}$. The following proposition establishes important statistical properties of the sequence $\{p_n(t)\}_{t \in \mathbb{T}_+}$:

Proposition 6

[i] For each n, the process $\{p_n(t)\}_{t \in \mathbb{T}_+}$ is a Markov chain on $V = [1, \cdots, N]$ with transition probability matrix $\overline{A}$.

[ii] The Markov chain $\{p_n(t)\}_{t \in \mathbb{T}_+}$ is ergodic with the uniform distribution on V being the attracting invariant measure.

Proof: For part [i], we note that, by the independence of $\{A(t)\}$, for any $t \in \mathbb{T}_+$ and $l_t, \cdots, l_0 \in V$,

$$
\begin{aligned}
\mathbb{P}\left[p_n(t) = l_t \,|\, p_n(t-1) = l_{t-1}, \cdots, p_n(1) = l_1, p_n(0) = l_0\right] &= \mathbb{P}\left[p_n(t) = l_t \,|\, p_n(t-1) = l_{t-1}\right] \\
&= \mathbb{P}\left[\to (l_{t-1}, t-1) = l_t\right] \\
&= \mathbb{P}\left[A_{l_{t-1}, l_t}(t-1) = 1\right] \\
&= \overline{A}_{l_{t-1}, l_t} \tag{14}
\end{aligned}
$$

where the last step follows from the fact, that the entries of $A(t-1)$ are binary. This establishes the desired Markovianity of the sequence $\{p_n(t)\}_{t \in \mathbb{T}_+}$.

For part [ii], since the state space V is finite, the irreducibility of $\overline{A}$ implies its positive recurrence and hence the invariant measure (the uniform distribution on V) is unique. That this measure is attracting follows from the aperiodicity of $\overline{A}$. ∎

For each of the Markov chains $\{p_n(t)\}_{t \in \mathbb{T}_+}$, we define a sequence of switched Riccati iterates $\{P_n(t)\}$:

$$P_n(t+1) = f_{p_n(t)}(P_n(t)) \tag{15}$$

The sequence $\{P_n(t)\}_{t \in \mathbb{T}_+}$ can be viewed as an iterated system of Riccati maps, the random switching sequence being governed by the Markov chain $\{p_n(t)\}_{t \in \mathbb{T}_+}$. A more intuitive explanation comes from the particle interpretation, precisely the n-th sequence may be viewed as a particle originating at node n and hopping around the network as a Markov chain with transition probability $\overline{A}$ whose instantaneous

state $P_n(t)$ evolves by the application of the Riccati operator corresponding to its current location. In particular, in contrary to the sequence of conditional error covariances at sensor n, $\left\{\widehat{P}_n(t)\right\}_{t\in\mathbb{T}_+}$, the sequence $\{P_n(t)\}_{t\in\mathbb{T}_+}$ does not correspond to the error evolution at a particular sensor. The following proposition shows that the sequence $\{P_n(t)\}_{t\in\mathbb{T}_+}$ is semi-Markov and establishes its relation to the sequence $\left\{\widehat{P}_n(t)\right\}_{t\in\mathbb{T}_+}$ of interest.

Proposition 7

[i] The sequence $\{P_n(t)\}_{t\in\mathbb{T}_+}$ is semi-Markov, given the Markov switching sequence, i.e.,

$$\mathbb{E}\left[\mathbb{I}_\Gamma\left(P_n(t+1)\right)\Big|\{P_n(s),p_n(s)\}_{0\le s\le t}\right] = \mathbb{E}\left[\mathbb{I}_\Gamma\left(P_n(t+1)\right)|P_n(t),p_n(t)\right], \quad \forall t\in\mathbb{T}_+, \ \ \Gamma\in\mathcal{B}(\mathbb{S}_+^N) \quad (16)$$

[ii] Consider the sequence of random permutations $\{\pi_t\}_{t\in\mathbb{T}_+}$ on V, given by

$$(\pi_{t+1}(1),\cdots,\pi_{t+1}(N)) = (\to(\pi_t(1),t),\cdots,\to(\pi_t(N),t)) \quad (17)$$

with initial condition

$$(\pi_0(1),\cdots,\pi_0(N)) = (1,\cdots,N) \quad (18)$$

(Note that $\pi_t(n) = p_n(t)$ for every n, where $p_n(t)$ is defined in eqn. (13).) Then, for $t\in\mathbb{T}_+$,

$$(P_1(t+1),\cdots,P_N(t+1)) = \left(\widehat{P}_{\pi_t(1)}(t+1),\cdots,\widehat{P}_{\pi_t(N)}(t+1)\right) \quad (19)$$

Part [ii] of the above proposition suggests that the asymptotics of the desired sequence $\left\{\widehat{P}_n(t)\right\}_{t\in\mathbb{T}_+}$ for every n can be obtained by studying the same for the sequences $\{P_n(t)\}_{t\in\mathbb{T}_+}$. Also, part [i] of Proposition 7 demonstrates the nice structure of the sequence $\{P_n(t)\}_{t\in\mathbb{T}_+}$. In the following, in particular, we will show that the sequences $\{P_n(t)\}_{t\in\mathbb{T}_+}$ reach consensus in a weak sense, which by part [ii] will establish weak consensus for the sequences $\left\{\widehat{P}_n(t)\right\}_{t\in\mathbb{T}_+}$ of interest. Hence, in the subsequent sections, we will study the sequences $\{P_n(t)\}_{t\in\mathbb{T}_+}$, rather than working directly on the sequences $\left\{\widehat{P}_n(t)\right\}_{t\in\mathbb{T}_+}$ of interest, which involve a much more complicated statistical dependence.

III. MAIN RESULTS

In this section, we present and discuss the main results of the paper under Assumptions **C.1, S.1, D.1**, see page 7. The first result does not directly concern the sequences $\left\{\widehat{P}_n(t)\right\}$ for $n = 1,\cdots,N$, but sets the stage for presenting the key result regarding the convergence of these sequences and is of independent interest.

Theorem 8 For a given $\overline{A}$, let $\{\widetilde{p}(t)\}_{t\in\mathbb{T}_+}$ be a stationary Markov chain on V with transition probability matrix $\overline{A}$, i.e., $\widetilde{p}(0)$ is distributed uniformly on V. Let ν be a probability measure on $\mathbb{S}_+^N$ and consider the random process $\left\{\widetilde{P}(t)\right\}_{t\in\mathbb{T}_+}$ given by

$$\widetilde{P}(t+1) = f_{\widetilde{p}(t)}\left(\widetilde{P}(t)\right), \quad t\in\mathbb{T}_+ \quad (20)$$

where $\widetilde{P}(0)$ is distributed as ν and independent of the Markov chain $\{\widetilde{p}(t)\}$. Then, there exists a probability measure (unique) $\mu^{\overline{A}}$ (depending on $\overline{A}$ only,) such that, for every ν, the process $\left\{\widetilde{P}(t)\right\}$ constructed above converges weakly to $\mu^{\overline{A}}$. In other words, for any ν if $\widetilde{P}(0) \sim \nu$ and independent of $\{\widetilde{p}(t)\}$, we have as $t \to \infty$ that the composition of Riccati operators converges in distribution

$$f_{\widetilde{p}(t)} \circ f_{\widetilde{p}(t-1)} \cdots \circ f_{\widetilde{p}(0)}\left(\widetilde{P}(0)\right) \Longrightarrow \mu^{\overline{A}} \tag{21}$$

Remark 9 We stress here that the dependence of the invariant measure $\mu^{\overline{A}}$ on the communication policy $\mathcal{D}$ manifests only through the mean matrix $\overline{A}$.

We now state the key result characterizing the convergence properties of the sequences $\left\{\widehat{P}_n(t)\right\}$.

Theorem 10 [i] Let q be a uniformly distributed random variable on V and independent of the sequence of adjacency matrices $\{A(t)\}_{t \in \mathbb{T}_+}$. Then, the sequence $\left\{\widehat{P}_q(t)\right\}_{t \in \mathbb{T}_+}$ converges weakly to $\mu^{\overline{A}}$ (the latter being defined in Theorem 8), i.e.,

$$\widehat{P}_q(t) \Longrightarrow \mu^{\overline{A}} \tag{22}$$

In other words, the conditional error covariance $\left\{\widehat{P}_q(t)\right\}$ of any randomly selected sensor (estimator) converges in distribution to $\mu^{\overline{A}}$.

[ii] For every $n \in [1, \cdots, N]$, the sequence $\{P_n(t)\}_{t \in \mathbb{T}_+}$ (or the sequence $\left\{\widehat{P}_{\pi_t(n)}(t)\right\}_{t \in \mathbb{T}_+}$ is stochastically dominated by the distribution $\mu^{\overline{A}}$ as $t \to \infty$, i.e., for every $\alpha > 0$, we have

$$\limsup_{t \to \infty} \mathbb{P}\left(\|P_n(t)\| \geq \alpha\right) \leq \mu^{\overline{A}}\left(\left\{X \in \mathbb{S}_+^N \,|\, \|X\| \geq \alpha\right\}\right) \tag{23}$$

$$\limsup_{t \to \infty} \mathbb{P}\left(P_n(t) \succeq \alpha I\right) \leq \mu^{\overline{A}}\left(\left\{X \in \mathbb{S}_+^N \,|\, X \succeq \alpha I\right\}\right) \tag{24}$$

More generally, for a closed set F preserving monotonicity, i.e., $X \in F$ implies $Y \in F$ for all $Y \succeq X$, we have

$$\limsup_{t \to \infty} \mathbb{P}\left(P_n(t) \in F\right) \leq \mu^{\overline{A}}\left(F\right) \tag{25}$$

In words, $\forall n$, the pathwise error associated with $\widehat{\mathbf{x}}_{\pi_t(n)}(t)$ is stochastically dominated by $\mu^{\overline{A}}$.

[iii] For each n, the sequence of error covariances $\left\{\widehat{P}_n(t)\right\}_{t \in \mathbb{T}_+}$ is stochastically bounded,

$$\lim_{J \to \infty} \sup_{t \in \mathbb{T}_+} \mathbb{P}\left(\left\|\widehat{P}_n(t)\right\| \geq J\right) = 0 \tag{26}$$

Specifically, for all closed F, we have

$$\limsup_{t \to \infty} \mathbb{P}\left(\widehat{P}_n(t) \in F\right) \leq N\mu^{\overline{A}}\left(F\right) \tag{27}$$

We discuss the consequences of Theorem 10. The first part of the theorem reinforces the weak consensus achieved by the GIKF algorithm, i.e., the conditional error covariance at a randomly selected sensor converges in distribution to the invariant measure $\mu^{\overline{A}}$. Reinterpreted, it provides an estimate $\{\widehat{\mathbf{x}}_q(t)\}$ (in

practice, obtained by uniformly selecting a sensor q independent of the random gossip protocol $\{A(t)\}$ and using its estimate $\widehat{\mathbf{x}}_q(t)$ for all time t) with stochastically bounded conditional error covariance under the weak detectability and connectivity assumptions. Note that the results provided in this paper pertain to the limiting distribution of the conditional error covariance and, hence, the pathwise filtering error. This is a much stronger result than providing moment estimates of the conditional error covariance, which does not provide much insight into the pathwise instantiations of the filter. In this paper, we do not provide analytic characterizations of the resulting invariant measure $\mu^{\overline{A}}$. However, Theorem 8 also provides an efficient numerical characterization of $\mu^{\overline{A}}$. In particular, the weak convergence in eqn. (21) shows that the empirical distribution obtained by plotting repeated instantiations of the process $\left\{\widetilde{P}(t)\right\}$ (eqn. (20)) would converge to $\mu^{\overline{A}}$.

Another class of estimates obtained by the GIKF algorithm is demonstrated in the second part of Theorem 10. For each n, the estimate $\left\{\widehat{\mathbf{x}}_{\pi_t(n)}(t)\right\}$ is obtained in practice by starting at the node n and then performing a random walk, $\pi_t(n)$, through the graph and collecting the estimates on the way. Eqns. (23-25) show that, in the limit as $t \to \infty$, these estimates are at least as good as the estimate $\{\widehat{\mathbf{x}}_q(t)\}$ obtained by probing a randomly selected node and using its estimate throughout. For some n, whether the estimate $\left\{\widehat{\mathbf{x}}_{\pi_t(n)}(t)\right\}$ is strictly better than the estimate $\{\widehat{\mathbf{x}}_q(t)\}$ asymptotically is an interesting technical question and not resolved in this paper. On the contrary, another possibility may be an extension of eqn. (25) to all closed F leading to the weak convergence of $\left\{\widehat{P}_{\pi_t(n)}(t)\right\}$ to $\mu^{\overline{A}}$ by Portmanteau's theorem. However, the inequality in eqn. (25) cannot be strict for all n, as we have for all closed F and $\varepsilon > 0$ (see Subsection VI-B,)

$$\frac{1}{N} \sum_{n=1}^{N} \liminf_{t \to \infty} \mathbb{P}\left(\widehat{P}_{\pi_t(n)}(t) \in F^\varepsilon\right) \geq \mu^{\overline{A}} \tag{28}$$

The last part of Theorem 10 shows that weak detectability (which is necessary for the error of a centralized estimator to be stochastic bounded) is sufficient in the distributed gossip setting to lead to sensor estimates with stochastically bounded errors. The upper bound presented in eqn. (27) is highly conservative and in fact, we have for all closed F (see Subsection VI-B,)

$$\sum_{n=1}^{N} \limsup_{t \to \infty} \mathbb{P}\left(\widehat{P}_n(t) \in F\right) \leq N\mu^{\overline{A}} \tag{29}$$

IV. THE AUXILIARY SEQUENCE $\{\widetilde{P}_t\}$: RDS FORMULATION

The asymptotic analysis of the semi-Markov processes $\{P_n(t)\}$ for $n = 1, \cdots, N$ does not fall under the purview of standard approaches based on iterated random systems ([25]) or a random dynamical system (RDS) ([26]) as the switching Markov chains $\{p_n(t)\}$ are non-stationary. In this section, we consider an auxiliary process $\left\{\widetilde{P}(t)\right\}$ whose evolution is governed by similar random Riccati iterates, the difference being that the switching Markov chain is stationary i.e., the switching Markov chain $\{\widetilde{p}(t)\}$ is initialized with the uniform invariant measure on V. We analyze the asymptotic properties of

the auxiliary sequence $\left\{\widetilde{P}(t)\right\}$ by formulating it as a RDS on the space $\mathbb{S}_+^N$ and then in subsequent sections we derive the asymptotics of the sequences $\{P_n(t)\}$ for $n = 1, \cdots, N$ through comparison arguments. We start by formally defining the sequence $\left\{\widetilde{P}(t)\right\}$[10]:

Consider a Markov chain on the graph V, $\{\widetilde{p}(t)\}_{t \in \mathbb{T}_+}$, with transition matrix $\overline{A}$ and uniform initial distribution, i.e.,

$$\mathbb{P}\left[\widetilde{p}(0) = n\right] = \frac{1}{N}, \quad n = 1, \cdots, N \tag{30}$$

By Proposition 6, the Markov chain $\{\widetilde{p}(t)\}$ is stationary.

We now define the auxiliary process $\left\{\widetilde{P}(t)\right\}$ as follows:

$$\widetilde{P}(t+1) = f_{\widetilde{p}(t)}\left(\widetilde{P}(t)\right) \tag{31}$$

with (possibly random) initial condition $\widetilde{P}(0)$.[11]

Before reading the next two Sections, we refer the reader to Appendix A where we review preliminary facts and results from the theory of monotone, sublinear random dynamical systems (RDS) ([27]) tailored to our needs. We then show in Subsection IV-A that the sequence $\left\{\widetilde{P}(t)\right\}$, for each n, admits an ergodic RDS formulation evolving on $\mathbb{S}_+^N$ and establish some of its properties in Subsection IV-B.

A. RDS formulation of $\left\{\widetilde{P}(t)\right\}$

In this subsection, we construct a RDS (θ^R, φ^R) on $\mathbb{S}_+^N$, which is equivalent to the auxiliary sequence $\left\{\widetilde{P}(t)\right\}$ in distribution. To this end, we construct the Markov chain $\{\widetilde{p}(t)\}$ (in a distributional sense) on the canonical path space. Let $\widetilde{\Omega}$ denote the set $\{1, \cdots, N\}$ with $\widetilde{\mathcal{F}}$ denoting the corresponding Borel algebra on $\widetilde{\Omega}$, which coincides with the power set of $\{1, \cdots, N\}$. Denote by Ω^R the two-sided infinite product of sets $\widetilde{\Omega}$, $\Omega^R = \bigotimes_{t=-\infty}^{\infty} \widetilde{\Omega}$, i.e., Ω^R is the space of double-sided sequences of entries in $\{1, \cdots, N\}$, i.e.,

$$\Omega^R = \{\omega = (\cdots, \omega_{-1}, \omega_0, \omega_1, \cdots) \,|\, \omega_t \in \{1, \cdots, N\}, \quad \forall t \in \mathbb{T}\} \tag{32}$$

We equip Ω^R with the corresponding product Borel algebra $\mathcal{F}^R = \bigotimes_{t=-\infty}^{\infty} \widetilde{\mathcal{F}}$ generated by the cylinder sets. Note that $\{\omega_t\}_{t \in \mathbb{T}_+}$ for all $\omega \in \Omega^R$ denotes the canonical path space (trajectory) of the Markov chain $\{\widetilde{p}(t)\}_{t \in \mathbb{T}_+}$. The reason for introducing two-sided sequences is a matter of technical convenience and will be evident soon. Consider the unique probability measure $\mathbb{P}^R$ on $\mathcal{F}^R$, under which the stochastic process (two-sided) $\{\omega_t\}_{t \in \mathbb{T}}$ is a stationary Markov chain on the finite state space $\{1, \cdots, N\}$ with transition probability matrix $\overline{A}$. By the assumption of stationarity and Proposition 6, the distribution of ω_t for each $t \in \mathbb{T}$ is necessarily the uniform distribution on $\{1, \cdots, N\}$. In particular, we note that the stochastic

---

[10]We are interested in the distributional properties of the various processes of concern. The actual pathwise construction is not of importance as long as the required distributional equivalence holds. We assume that the measure space $(\Omega, \mathcal{F}, \mathbb{P})$ is rich enough (or suitably extended) to carry out constructions of the various auxiliary random variables.

[11]Although the sequences $\{P_n(t)\}$ of interest have deterministic initial conditions, it is required for technical reasons (to be made precise later) to allow random initial states $\widetilde{P}(0)$, when studying the auxiliary sequence $\left\{\widetilde{P}(t)\right\}$.

processes $\{\widetilde{p}(t)\}_{t\in\mathbb{T}}$ and $\{\omega_t\}_{t\in\mathbb{T}}$ are equivalent in terms of the distribution induced on path space. Define the family of transformations $\left\{\theta_t^R\right\}_{t\in\mathbb{T}}$ on Ω as the family of left-shifts, i.e,

$$\theta_t^R\omega = \omega(t + \cdot), \quad \forall t \in \mathbb{T} \tag{33}$$

With this, the space $\left(\Omega^R, \mathcal{F}^R, \mathbb{P}^R, \{\theta_t^R, t \in \mathbb{T}\}\right)$ becomes the canonical path space of a two-sided stationary sequence equipped with the left-shift operator and hence (see, for example, [28]) satisfies the Assumptions **A.1)**-**A.3)** in Definition 17 to be a metric dynamical system and, in fact, is also ergodic.

We now set to define the cocycle φ^R, see also Definition 17, over $\mathbb{S}_+^N$, which gives the RDS of interest. We define $\varphi^R : \mathbb{T}_+ \times \Omega^R \times \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ by:

$$\varphi^R(0, \omega, X) = X, \quad \forall \omega, X \tag{34}$$

$$\varphi^R(1, \omega, X) = f_{\omega_0}(X), \quad \forall \omega, X \tag{35}$$

$$\varphi^R(t, \omega, X) = f_{\theta_{t-1}^R\omega(0)}\left(\varphi^R(t-1, \omega, X)\right) = f_{\omega_{t-1}}\left(\varphi^R(t-1, \omega, X)\right), \quad \forall t > 1, \omega, X \tag{36}$$

(Note that, by property of the left shift θ^R, we have $\theta_{t-1}^R\omega(0) = \omega_t$, which explains the equality in eqn. (36).) The cocycle φ^R defined satisfies the assumptions of measurability jointly in its arguments, and the continuity of the map $\varphi^R(t, \omega, \cdot) : \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ w.r.t. the phase variable X for each fixed t, ω follows from the continuity of the corresponding Riccati operator. The pair $\left(\theta^R, \varphi^R\right)$ thus forms a well-defined RDS on the phase space $\mathbb{S}_+^N$. Now consider the sequence of random variables $\left\{\varphi^R(t, \omega, P_n(0))\right\}_{t\in\mathbb{T}_+}$ (as explained earlier, the randomness is induced by ω,) which can be viewed as successive (random) iterates of the RDS $\left(\theta^R, \varphi^R\right)$ starting with the initial state $P_n(0)$. By construction, it follows that the sequence $\left\{\varphi^R(t, \omega, P_n(0))\right\}_{t\in\mathbb{T}_+}$ is distributionally equivalent to the sequence $\left\{\widetilde{P}(t)\right\}_{t\in\mathbb{T}_+}$. In particular,

$$\varphi^R\left(t, \omega, P_n(0)\right) \overset{d}{=} \widetilde{P}(t), \quad \forall t \in \mathbb{T}_+ \tag{37}$$

Thus, analyzing the asymptotic distributional properties of the sequence $\left\{\widetilde{P}(t)\right\}_{t\in\mathbb{T}_+}$ is equivalent to studying the sequence $\left\{\varphi^R\left(t, \omega, P_n(0)\right)\right\}_{t\in\mathbb{T}_+}$, which we undertake in the next subsection.

B. Properties of the RDS (θ^R, φ^R)

We establish some basic properties of the RDS $\left(\theta^R, \varphi^R\right)$ representing the auxiliary sequence $\left\{\widetilde{P}(t)\right\}$.

Lemma 11

[i] The RDS $\left(\theta^R, \varphi^R\right)$ is conditionally compact.

[ii] The RDS $\left(\theta^R, \varphi^R\right)$ is order preserving.

[iii] If in addition $\mathcal{Q}$ is positive definite, i.e., $\mathcal{Q} \gg 0$, then $\left(\theta^R, \varphi^R\right)$ is strongly sublinear.

Proof: The claim in [i] (conditional compactness) is an immediate consequence of the finite dimensionality of the underlying vector space $\mathbb{S}_+^N$.

The order preserving property [ii] follows from the monotonicity of the individual Riccati operators f_n and hence finite compositions of them remain order-preserving.

The strong sublinearity uses the concavity of the Riccati operators and their monotone nature and is a routine extension to an arbitrary number N of Riccati operators, given the development in [19] (see Lemma 21 in [19]) for the case of two Riccati operators. ∎

V. ASYMPTOTICS OF $\left\{ \widetilde{P}(t) \right\}$

The main result here concerns the asymptotic properties of the auxiliary sequences $\left\{ \widetilde{P}(t) \right\}_{t \in \mathbb{T}_+}$ for each $n \in [1, \cdots, N]$. We have the following:

Theorem 12 Under the assumptions **C.1,S.1,D.1**, see page 7, there exists a unique equilibrium probability measure $\mu^{\overline{A}}$ on the space of positive semidefinite matrices $\mathbb{S}_+^N$, such that, for each $n \in [1, \cdots, N]$, the sequence $\left\{ \widetilde{P}(t) \right\}_{t \in \mathbb{T}_+}$ converges weakly (in distribution) to $\mu^{\overline{A}}$ from every initial condition $P_n(0)$:

$$\left\{ \widetilde{P}(t) \right\} \Longrightarrow \mu^{\overline{A}}, \quad \forall n \in [1, \cdots, N] \tag{38}$$

The rest of the subsection is devoted to the proof of the above result. But, before that, we highlight some consequences of Theorem 12.

Remark 13 It is important to note, as stated in Theorem 12, that the equilibrium measure $\mu^{\overline{A}}$ does not depend on the index n and the initial state $\widetilde{P}(0)$ of the sequence $\left\{ \widetilde{P}(t) \right\}$, but is a functional of the network topology and the particular (randomized) communication protocol captured by the matrix $\overline{A}$. Theorem 12, thus concludes that the sequences $\left\{ \widetilde{P}(t) \right\}$ reach consensus *in the weak sense* to the same equilibrium measure irrespective of the initial states.

The proof of Theorem 12 is rather long and technical, which we accomplish in steps.

Lemma 14 Recall Assumption **D.1**, page 7, and let, in particular, $w_0 = (n_1, \cdots, n_\ell)$ be a walk such that, the Grammian

$$G_{w_0} = \sum_{i=1}^{\ell} \left(\mathcal{F}^{i-1} \right)^T \mathcal{C}_{n_i}^T \mathcal{C}_{n_i} \mathcal{F}^{i-1} \tag{39}$$

is invertible, where $\ell \geq 1$. Define the function $g_{w_0} : \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ by

$$g_{w_0}(X) = f_{n_\ell} \circ f_{n_\ell - 1} \circ \cdots \circ f_{n_1}(X) \tag{40}$$

Then, there exists a constant $\alpha_0 > 0$ such that the following uniformity condition holds:

$$g_{w_0}(X) \leq \alpha_0 I, \quad \forall X \in \mathbb{S}_+^N \tag{41}$$

In other words the iterate $g_{w_0}(\cdot)$ is uniformly bounded irrespective of the value of the argument.

The proof is provided in Appendix B. Note that in eqn. (39) the observation matrix $\mathcal{C}_{n_i}$ is indexed by n_i the current site visited by the random walk w_0 introduced in Lemma 14. Also, note that the function $g_{w_0}(X)$ defined in eqn. (40) is indexed by the walk w_0.

The following key lemma establishes asymptotic boundedness properties of $\left\{\widetilde{P}(t)\right\}$ and is proved in Appendix B.

Lemma 15 The sequence $\left\{\widetilde{P}(t)\right\}$ is stochastically bounded for each n under the Assumptions of Theorem 12, i.e.,

$$\lim_{J \to \infty} \sup_{t \in \mathbb{T}_+} \mathbb{P}\left(\left\|\widetilde{P}(t)\right\| > J\right) = 0 \tag{42}$$

We now complete the proof of Theorem 12.

From Lemma IV-A we note that $\left(\theta^R, \varphi^R\right)$ is strongly sublinear, conditionally compact and order-preserving. Also, the cone $\mathbb{S}_+^N$ satisfies the conditions required in the hypothesis of Theorem 26. We note for $t > 0$

$$\varphi^R(t, \omega, 0) = f_{\omega(t-1)}\left(\varphi(t-1, \omega, 0)\right) \succeq \mathcal{Q} \gg 0 \tag{43}$$

Thus the hypotheses of Theorem 26 are satisfied, and precisely one of the assertions **a)** and **b)** holds. By an argument similar to Lemma 23 in [19], we can show that assertion **a)** cannot hold in the face of stochastic boundedness of the sequence $\left\{\widetilde{P}(t)\right\}_{t \in \mathbb{T}_+}$ (Lemma 15). Thus assertion **b)** holds, and, as a direct consequence of Theorem 26, we establish the existence of a unique almost equilibrium $u^{\overline{A}}(\omega) \gg 0$ defined on a θ^R-invariant set $\Omega^* \in \mathcal{F}^R$ with $\mathbb{P}\left(\Omega^*\right) = 1$ such that, for any random variable $v(\omega)$ possessing the property $0 \preceq v(\omega) \preceq \alpha u^{\overline{A}}(\omega)$ for all $\omega \in \Omega^*$ and deterministic $\alpha > 0$, the following holds:

$$\lim_{t \to \infty} \varphi\left(t, \theta_{-t}\omega, v(\theta_{-t}\omega)\right) = u^{\overline{A}}(\omega), \quad \omega \in \Omega^* \tag{44}$$

From the distributional equivalence of pull-back and forward orbits, Lemma 23 establishes the existence of a unique almost equilibrium $u^{\overline{A}}$, i.e., a unique equilibrium measure for the process $\left\{\widetilde{P}(t)\right\}$ from the distributional equivalence of pull-back and forward orbits. However, to show that the measure induced by $u^{\overline{A}}$ on $\mathbb{S}_+^N$ is attracting for $\left\{\widetilde{P}(t)\right\}$, eqn. (44) must hold for all initial v, whereas Lemma 23 establishes convergence for a restricted class of initial conditions v. We need the following result to extend it to general initial conditions.

Lemma 16 Under the assumptions of Theorem 12, let $u^{\overline{A}}$ be the unique almost equilibrium of the RDS $\left(\theta^R, \varphi^R\right)$. Then

$$\mathbb{P}\left(\omega : u^{\overline{A}}(\omega) \succeq \mathcal{Q}\right) = 1 \tag{45}$$

Proof: The proof uses the fact that, for all n, $f_n(X) \succeq \mathcal{Q}$, and is routine given the corresponding development in Lemma 24 of [19]. ∎

We now complete the proof of Theorem 12.

Proof of Theorem 12: : Let $\mu^{\overline{A}}$ be the distribution of the unique almost equilibrium in eqn. (44). By Lemma 16 we have $\mu^{\overline{A}}\left(\mathbb{S}_{++}^N\right) = 1$. Let $P_0 \in \mathbb{S}_+^N$ be an arbitrary initial state. By construction of the RDS $\left(\theta^R, \varphi^R\right)$, the sequences $\{P_t\}_{t \in \mathbb{T}_+}$ and $\left\{\varphi^R(t, \omega, P_0)\right\}_{t \in \mathbb{T}_+}$ are distributionally equivalent, i.e., $P_t \overset{d}{=} \varphi^R(t, \omega, P_0)$. Recall Ω^* as the θ^R-invariant set with $\mathbb{P}^{\overline{\gamma}}\left(\Omega^*\right) = 1$ in eqn. (44) on which the almost

equilibrium $u^{\overline{A}}$ is defined. By Lemma 16, there exists $\Omega_1 \subset \Omega^*$ with $\mathbb{P}^{\overline{\gamma}}(\Omega_1) = 1$, such that

$$u^{\overline{A}}(\omega) \succeq \mathcal{Q}, \quad \omega \in \Omega_1 \tag{46}$$

Define the random variable $\widetilde{X} : \Omega \longmapsto \mathbb{S}_+^N$ by

$$\begin{cases} P_0 & \text{if } \omega \in \Omega_1 \\ 0 & \text{if } \omega \in \Omega_1^c \end{cases} \tag{47}$$

Now choose $\alpha > 0$ sufficiently large such that

$$P_0 \preceq \alpha \mathcal{Q} \tag{48}$$

This is possible because $\mathcal{Q} \gg 0$. Then

$$0 \preceq \widetilde{X}(\omega) \preceq \alpha u^{\overline{A}}(\omega), \quad \omega \in \Omega^* \tag{49}$$

Indeed, we have

$$0 \preceq P_0 = \widetilde{X}(\omega) \preceq \alpha \mathcal{Q} \preceq \alpha u^{\overline{A}}(\omega), \quad \omega \in \Omega_1 \tag{50}$$

and

$$0 = \widetilde{X}(\omega) \preceq \alpha u^{\overline{A}}(\omega), \quad \omega \in \Omega \backslash \Omega_1 \tag{51}$$

We then have by the discussion preceding eqn. (44)

$$\lim_{t \to \infty} \varphi^R \left(t, \theta_{-t}\omega, \widetilde{X}(\theta_{-t}\omega) \right) = u^{\overline{A}}(\omega), \quad \omega \in \Omega^* \tag{52}$$

Since convergence $\mathbb{P}^{\overline{\gamma}}$ a.s. implies convergence in distribution, we have

$$\varphi^R \left(t, \theta_{-t}\omega, \widetilde{X}(\theta_{-t}\omega) \right) \Longrightarrow \mu^{\overline{A}} \tag{53}$$

as $t \to \infty$, where $\Longrightarrow$ denotes weak convergence or convergence in distribution. Then, by Lemma 23, the sequence $\left\{ \varphi^R \left(t, \omega, \widetilde{X}(\omega) \right) \right\}_{t \in \mathbb{T}_+}$ also converges in distribution to the unique stationary distribution $\mu^{\overline{A}}$, i.e., as $t \to \infty$

$$\varphi^R \left(t, \omega, \widetilde{X}(\omega) \right) \Longrightarrow \mu^{\overline{A}} \tag{54}$$

Now, since $\mathbb{P}^{\overline{\gamma}}(\Omega_1) = 1$, by eqn. (47)

$$\varphi^R (t, \omega, P_0) = \varphi^R \left(t, \omega, \widetilde{X}(\omega) \right), \quad \mathbb{P}^{\overline{\gamma}} \text{ a.s.}, \ t \in \mathbb{T}_+ \tag{55}$$

which implies

$$\varphi^R (t, \omega, P_0) \stackrel{d}{=} \varphi^R \left(t, \omega, \widetilde{X}(\omega) \right), \quad t \in \mathbb{T}_+ \tag{56}$$

From eqns. (54,56), we then have $\varphi^R (t, \omega, P_0) \Longrightarrow \mu^{\overline{A}}$, which together with the distributional equivalence $P_t \stackrel{d}{=} \varphi^R (t, \omega, P_0)$ noted above implies, as $t \to \infty$, $P_t \Longrightarrow \mu^{\overline{A}}$. $\blacksquare$

VI. Proofs of main results

A. Proof of Theorem 8

Proof: By Theorem 12 we know that such a sequence $\left\{\widetilde{P}(t)\right\}$ converges weakly to $\mu^{\overline{A}}$ when started from a deterministic initial condition. In the case, $\widetilde{P}(0)$ is distributed as ν, we note that, by the independence of $\widetilde{P}(0)$ and the Markov chain $\{q(t)\}$,

$$\mathbb{E}\left[g\left(\widetilde{P}(t)\right)\right] = \int_{\mathbb{S}_+^N} \mathbb{E}\left[\left(\widetilde{P}(t)\right)\Big|\widetilde{P}(0) = X\right] d\nu(X) \tag{57}$$

for any $g \in C_b(\mathbb{S}_+^N)$. Now, the distribution of the sequence $\left\{\widetilde{P}(t)\right\}$ conditioned on the event $\widetilde{P}(0) = X$ is the same as that when the sequence starts with the deterministic initial condition X (this is true because $\widetilde{P}(0)$ is independent of $\{q(t)\}$.) Hence by Theorem 12

$$\lim_{t\to\infty} \mathbb{E}\left[\left(\widetilde{P}(t)\right)\Big|\widetilde{P}(0) = X\right] = \int_{\mathbb{S}_+^N} g(y)d\mu^{\overline{A}}(Y) \tag{58}$$

for all X. Since g is bounded, the dominated convergence theorem and eqn. (57) lead to

$$\lim_{t\to\infty} \mathbb{E}\left[\left(\widetilde{P}(t)\right)\right] = \int_{\mathbb{S}_+^N} g(y)d\mu^{\overline{A}}(Y) \tag{59}$$

for all $g \in C_b(X)$, and hence the required weak convergence follows. ∎

B. Proof of Theorem 10

Proof: We prove Theorem 10 in the order **1),3)** and **2)**.

Consider any $\Gamma \in \mathcal{B}(\mathbb{S}_+^N)$. We estimate the probability $\mathbb{P}\left(\widehat{P}_q(t) \in \Gamma\right)$. To this end, we note that

$$\mathbb{P}\left(\widehat{P}_q(t) \in \Gamma\right) = \sum_{n=1}^N \mathbb{P}\left(\widehat{P}_n(t) \in \Gamma\right)\mathbb{P}(q = n) = \frac{1}{N}\sum_{n=1}^N \mathbb{P}\left(\widehat{P}_n(t) \in \Gamma\right) \tag{60}$$

The first step holds because q is independent of the sequences $\left\{\widehat{P}_n(t)\right\}$ for all n and subsequently we use that q is uniformly distributed on V. Denoting by π_t^{-1} the inverse of the permutation π_t, we have

$$\mathbb{P}\left(\widehat{P}_n(t) \in \Gamma\right) = \mathbb{P}\left(P_{\pi_t^{-1}(n)}(t) \in \Gamma\right) = \sum_{l=1}^N \mathbb{P}\left(\{P_l(t) \in \Gamma\}\bigcap\{\pi_t^{-1}(n) = l\}\right) \tag{61}$$

Note, here, unlike in eqn. (60), we may not gain much by splitting the probabilities in the last step as the events $\{P_l(t) \in \Gamma\}$ and $\{\pi_t^{-1}(n) = l\}$ are not independent. Combining eqns. (60,61), we have

$$\begin{aligned}
\mathbb{P}\left(\widehat{P}_q(t) \in \Gamma\right) &= \frac{1}{N}\sum_{n=1}^N\sum_{l=1}^N \mathbb{P}\left(\{P_l(t) \in \Gamma\}\bigcap\{\pi_t^{-1}(n) = l\}\right) \\
&= \frac{1}{N}\sum_{l=1}^N\sum_{n=1}^N \mathbb{P}\left(\{P_l(t) \in \Gamma\}\bigcap\{\pi_t^{-1}(n) = l\}\right) \\
&= \frac{1}{N}\sum_{l=1}^N \mathbb{P}\left(P_l(t) \in \Gamma\right)
\end{aligned} \tag{62}$$

Note the last step follows from the fact that

$$\sum_{n=1}^{N} \mathbb{P}\left(\{P_l(t) \in \Gamma\} \bigcap \{\pi_t^{-1}(n) = l\}\right) = \mathbb{P}\left(P_l(t) \in \Gamma\right) \tag{63}$$

because the events $\{\pi_t^{-1}(n) = l\}$, $n = 1, \cdots, N$ are mutually exclusive and exhaustive, $\pi^{-1}(t)$ being a permutation.

Now consider a stationary Markov chain $\{\widetilde{p}(t)\}$ on V with transition probability $\overline{A}$ and let $\left\{\widetilde{P}(t)\right\}$ be the sequence defined by

$$\widetilde{P}(t+1) = f_{\widetilde{p}(t)}\left(\widetilde{P}(t)\right), \quad t \in \mathbb{T}_+ \tag{64}$$

with initial condition $\widetilde{P}(0) = \widehat{P}(0)$. Then,

$$\mathbb{P}\left(\widetilde{P}(t) \in \Gamma\right) = \sum_{l=1}^{N} \mathbb{P}\left(\widetilde{P}(t) \in \Gamma \,\middle|\, \widetilde{p}(0) = l\right) \mathbb{P}\left(\widetilde{p}(0) = l\right) = \frac{1}{N}\mathbb{P}\left(\widetilde{P}(t) \in \Gamma \,\middle|\, \widetilde{p}(0) = l\right) \tag{65}$$

By construction, the distribution of the sequence $\left\{\widetilde{P}(t)\right\}$ conditioned on the event $\{\widetilde{p}(0) = l\}$ is equivalent to that of the sequence $\{P_l(t)\}$ and hence

$$\mathbb{P}\left(\widetilde{P}(t) \in \Gamma \,\middle|\, \widetilde{p}(0) = l\right) = \mathbb{P}\left(P_l(t) \in \Gamma\right) \tag{66}$$

Hence by eqns. (62,66) we obtain

$$\mathbb{P}\left(\widehat{P}_q(t) \in \Gamma\right) = \mathbb{P}\left(\widetilde{P}(t) \in \Gamma\right) \tag{67}$$

Thus, for all t, $P_q(t) \stackrel{d}{=} \widetilde{P}(t)$. By Theorem 12, we then have the weak convergence of the sequence $\{P_q(t)\}$ to $\mu^{\overline{A}}$.

For the third part, we note that for any $\Gamma \in \mathbb{B}(\mathbb{S}_+^N)$

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{P}\left(\widehat{P}_n(t) \in \Gamma\right) = \mathbb{P}\left(\widehat{P}_q(t) \in \Gamma\right) \tag{68}$$

due to the independence of q from $\{A(t)\}$. Taking the $\limsup$ and noting the non-negativity of the terms, we have for closed F,

$$\begin{aligned}
\limsup_{t\to\infty} \mathbb{P}\left(\widehat{P}_n(t) \in F\right) &\leq \limsup_{t\to\infty} \sum_{n=1}^{N} \mathbb{P}\left(\widehat{P}_n(t) \in F\right) \\
&= N\limsup_{t\to\infty} \sum_{n=1}^{N} \left[\frac{1}{N}\mathbb{P}\left(\widehat{P}_n(t) \in F\right)\right] \\
&= N\limsup_{t\to\infty} \mathbb{P}\left(\widehat{P}_q(t) \in F\right) \leq N\mu^{\overline{A}}
\end{aligned}$$

The proof of the second part involves an auxiliary construction and approximation arguments to relate the limit properties of the sequences $\{P_n(t)\}$ to similar processes, where the underlying switching Markov chain is stationary. To this end consider any strictly positive $s \in \mathbb{T}_+$. Recall the Markov chains $\{p_n(t)\}$

for $n = 1, \cdots, N$ with transition probability matrix $\overline{A}$ and initial state $p_n(0) = n$. The corresponding sequence of interacting particle processes $\{P_n(t)\}$ are constructed, for each n as:

$$P_n(t+1) = f_{p_n(t)}(P_n(t)) \tag{69}$$

with initial condition $P_n(0) = \widehat{P}(0)$. Let $f_0 : \mathbb{S}_+^N \longmapsto \mathbb{S}_+^N$ denote the Lyapunov operator

$$f_0(X) = \mathcal{F} X \mathcal{F}^T + \mathcal{Q} \tag{70}$$

and note that the following ordering holds:

$$f_n(X) \preceq f_0(X), \quad \forall n \text{ and } X \in \mathbb{S}_+^N \tag{71}$$

For a given $s > 0$ chosen above and for all n, define the processes $\{P_n^s(t)\}_{t \geq s}$ by

$$P_n^s(t+1) = f_{p_n(t)}(P_n^s(t)) \tag{72}$$

with deterministic initial value $P_n^s(s) = f_0^s\left(\widehat{P}(0)\right)$. By eqn. (71), for any s tuple $(i_0, i_1, \cdots, i_{s-1})$ with $i_r \in [1, \cdots, N]$ for $r = 0, \cdots, s-1$, we note that

$$f_{i_{s-1}} \circ f_{i_{s-2}} \circ \cdots \circ f_{i_0}\left(\widehat{P}(0)\right) \preceq f_0^s\left(\widehat{P}(0)\right) \tag{73}$$

and, hence, by the monotonicity of the Riccati operators, we conclude that for all n

$$P_n(t) \preceq P_n^s(t), \quad t \geq s \tag{74}$$

Also consider a stationary Markov chain $\{q(t)\}_{t \geq s}$ with transition probability $\overline{A}$, i.e., $q(0)$ is uniformly distributed on V, and define the process $\{Q^s(t)\}_{t \geq s}$ by

$$Q^s(t+1) = f_{q(t)}(Q^s(t)) \tag{75}$$

with deterministic initial value

$$Q^s(s) = f_0^s\left(\widehat{P}(0)\right) \tag{76}$$

It is to be noted that by Theorem 12, the process $\{Q^s(t)\}$ converges weakly to $\mu^{\overline{A}}$, i.e.,

$$\lim_{t \to \infty} d_P\left(Q^s(t), \mu^{\overline{A}}\right) = 0 \tag{77}$$

where d_P denotes the Prohorov metric. We now set to relate the limit properties of $\{P_n^s(t)\}$ to those of $\{Q^s(t)\}$. For $t \geq s$ define the total variation distance between $P_n^s(t)$ and $Q^s(t)$ by

$$d_v\left(P_n^s(t), Q^s(t)\right) = \sup_{\Gamma \in \mathcal{B}(\mathbb{S}_+^N)} \left|\mathbb{P}\left(P_n^s(t) \in \Gamma\right) - \mathbb{P}\left(Q^s(t) \in \Gamma\right)\right| \tag{78}$$

Since for any t, the two sequences considered above assume values in a finite set, we define a set of

$(t - s)$ tuples $\Lambda(\Gamma)$ by

$$\Lambda(\Gamma) = \left\{ (i_1, \cdots, i_{t-s}) \mid i_r \in [1, \cdots, N] \text{ for all r and } f_{i_{t-s}} \circ \cdots \circ f_{i_1} \left(f_0^s(\widehat{P}(0)) \right) \right\} \tag{79}$$

It is clear that

$$\{P_n^s(t) \in \Gamma\} \iff \{(p_n(s), \cdots, p_n(t-1)) \in \Lambda(\Gamma)\} \tag{80}$$

$$\{Q^s(t) \in \Gamma\} \iff \{(q(s), \cdots, q(t-1)) \in \Lambda(\Gamma)\} \tag{81}$$

We then have

$$\begin{aligned}
\mathbb{P}\left(P_n^s(t) \in \Gamma\right) - \mathbb{P}\left(Q^s(t) \in \Gamma\right) &= \sum_{i_1} \mathbb{P}\left(p_n(s) = i_1\right) \sum_{(i_1, \cdots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \overline{A}_{i_r i_{r+1}} \\
&\quad - \sum_{i_1} \mathbb{P}\left(q(s) = i_1\right) \sum_{(i_1, \cdots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \overline{A}_{i_r i_{r+1}} \\
&= \sum_{i_1} \left[\mathbb{P}\left(p_n(s) = i_1\right) - \mathbb{P}\left(q(s) = i_1\right)\right] \sum_{(i_1, \cdots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \overline{A}_{i_r i_{r+1}}
\end{aligned}$$

and hence

$$\begin{aligned}
\left|\mathbb{P}\left(P_n^s(t) \in \Gamma\right) - \mathbb{P}\left(Q^s(t) \in \Gamma\right)\right| &\leq \sum_{i_1} \left|\mathbb{P}\left(p_n(s) = i_1\right) - \mathbb{P}\left(q(s) = i_1\right)\right| \sum_{(i_1, \cdots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \overline{A}_{i_r i_{r+1}} \\
&\leq \sum_{i_1} \left|\mathbb{P}\left(p_n(s) = i_1\right) - \mathbb{P}\left(q(s) = i_1\right)\right| \\
&\leq \sum_{i_1} d_v(p_n(s), q(s)) \leq N d_v\left(p_n(s), q(s)\right)
\end{aligned}$$

where we have used the fact that

$$\sum_{(i_1, \cdots, i_{t-s}) \in \Lambda(\Gamma)} \prod_{r=1}^{t-s-1} \overline{A}_{i_r i_{r+1}} = \mathbb{P}\left((p_n(s+1), \cdots, p_n(t-1)) = (i_2, \cdots, i_{t-s}) \mid p_n(s) = i_1\right) \leq 1$$

We thus obtain

$$d_v\left(P_s^n(t), Q^s(t)\right) \leq N d_v\left(p_n(s), q(s)\right), \quad \forall t \geq s \tag{82}$$

It is well known that the finite state Markov chain $\{p_n(s)\}$ converges weakly at a geometric rate to the uniform measure, i.e., the measure induced by $q(s)$ for each s and hence in variation. In other words,

$$\lim_{s \to \infty} d_v(p_n(s), q(s)) = 0 \tag{83}$$

Thus, by eqn. (82), we have

$$\lim_{s \to \infty} \sup_{t \geq s} d_v\left(P_s^n(t), Q^s(t)\right) = 0 \tag{84}$$

and, since convergence in total variation implies weak convergence ([29]), we have

$$\lim_{s \to \infty} \sup_{t \geq s} d_P \left(P_s^n(t), Q^s(t) \right) = 0 \tag{85}$$

Now consider $\varepsilon > 0$. Then there exists $s(\varepsilon)$, such that,

$$d_P \left(P_s^n(t), Q^s(t) \right) \leq \varepsilon/2, \quad s \geq s(\varepsilon), t \geq s \tag{86}$$

Since the sequence $\{Q^s(t)\}$ converges weakly to $\mu^{\overline{A}}$ for all s (in particular for $s = s(\varepsilon)$,) there exists $t(\varepsilon) \geq s(\varepsilon)$ sufficiently large, such that,

$$d_P \left(Q^{s(\varepsilon)}(t), \mu^{\overline{A}} \right) \leq \varepsilon/2, \quad t \geq t(\varepsilon) \tag{87}$$

Then, an application of the triangle inequality for the metric d_P leads to

$$d_P \left(P_n^{s(\varepsilon)}(t), \mu^{\overline{A}} \right) \leq d_P \left(P_n^{s(\varepsilon)}(t), Q^{s(\varepsilon)}(t) \right) + d_P \left(Q^{s(\varepsilon)}(t), \mu^{\overline{A}} \right) \leq \varepsilon \tag{88}$$

for all $t \geq t(\varepsilon)$. Now, by definition,

$$d_P \left(P_n^{s(\varepsilon)}(t), \mu^{\overline{A}} \right) = \inf \left\{ \delta > 0 \, \middle| \, \mathbb{P} \left(P_n^{s(\varepsilon)}(t) \in F \right) \leq \mu^{\overline{A}} \left(F^\delta \right) + \delta \text{ for all closed } F \in \mathbb{S}_+^N \right\} \tag{89}$$

where F^δ is defined as

$$F^\delta = \left\{ X \in \mathbb{S}_+^N \, \middle| \, \inf_{Y \in F} \|X - Y\| < \delta \right\} \tag{90}$$

Since, by eqn. (88), $d_P \left(P_n^{s(\varepsilon)}(t), \mu^{\overline{A}} \right) \leq \varepsilon$ for all $t \geq t(\varepsilon)$, we have, for any closed set F,

$$\mathbb{P} \left(P_n^{s(\varepsilon)}(t) \in F \right) \leq \mu^{\overline{A}} \left(F^\varepsilon \right) + \varepsilon, \quad t \geq t(\varepsilon) \tag{91}$$

In addition to F being closed, let us assume that F satisfies monotonicity, i.e., $X \in F$ implies $Y \in F$ for all $Y \succeq X$. By eqn. (74) we have

$$P_n(t) \preceq P_n^{s(\varepsilon)}(t), \quad t \geq t(\varepsilon) \geq s(\varepsilon) \tag{92}$$

and hence

$$\mathbb{P} \left(P_n(t) \in F \right) \leq \mathbb{P} \left(P_n^{s(\varepsilon)}(t) \in F \right), \quad t \geq t(\varepsilon) \tag{93}$$

We then have from eqn. (91) for all $t \geq t(\varepsilon)$

$$\mathbb{P} \left(P_n(t) \in F \right) \leq \mu^{\overline{A}} \left(F^\varepsilon \right) + \varepsilon \tag{94}$$

Taking the limit as $t \to \infty$, we have

$$\limsup_{t \to \infty} \mathbb{P} \left(P_n(t) \in F \right) \leq \mu^{\overline{A}} \left(F^\varepsilon \right) + \varepsilon \tag{95}$$

The L.H.S. above is now independent of t and, hence, ε through $t(\varepsilon)$. Since the above holds for arbitrary

$\varepsilon > 0$, moving to the limit as $\varepsilon \to 0$ yields

$$\limsup_{t \to \infty} \mathbb{P}\left(P_n(t) \in F\right) \leq \lim_{\varepsilon \to 0} \mu^{\overline{A}}\left(F^\varepsilon\right) = \mu^{\overline{A}}\left(F\right) \tag{96}$$

The last step follows from the continuity of the probability measure $\mu^{\overline{A}}$ and the fact that

$$\bigcap_{\varepsilon > 0} F^\varepsilon = F \tag{97}$$

for closed F. This establishes the result for general order preserving F. The result for sets of the form $\left\{X \in \mathbb{S}_+^N \,\middle|\, X \succeq \alpha I\right\}$ or $\left\{X \in \mathbb{S}_+^N \,\middle|\, \|X\| \geq \alpha\right\}$ for $\alpha > 0$ follow, as they satisfy the general hypothesis on F. ∎

VII. CONCLUDING REMARKS

The paper develops the gossip interactive Kalman filter (GIKF) for distributed Kalman filtering in sensor networks, when observation sampling and inter-sensor communication occur at the same time scale. Inter-sensor collaboration is achieved by intermittent exchange of filtering states. A traveling particle interpretation of the filtering states leads to a random dynamical system (RDS) formulation of the sequence of conditional error covariances. Under a weak detectability assumption, the estimation error process at each sensor stays stochastically bounded (irrespective of the instability in signal dynamics,) provided the network satisfies some weak connectivity conditions. Also, the network achieves weak consensus, i.e., the conditional error covariance (or the pathwise filtering error) at a randomly selected sensor converges in distribution to a unique invariant measure $\mu^{\overline{A}}$. The invariant measure $\mu^{\overline{A}}$ depends on the network connectivity process (the MAC protocol) through the mean $\overline{A}$ of the random adjacency matrix A.

The characterization of the invariant measure $\mu^{\overline{A}}$ as a functional of the matrix is interesting to study the sensitivity of the mapping, $\overline{A} \longrightarrow \mu^{\overline{A}}$. This would lead to understanding the robustness of the above filtering approach to perturbations in the communication policy, i.e., whether a small change in the MAC protocol (a perturbation of $\overline{A}$) leads to a negligible change of $\mu^{\overline{A}}$, or the filtering performance changes dramatically. Exploring such comparison principles for the mapping would lead to understanding the more complicated problem of characterizing the invariant measure $\mu^{\overline{A}}$. Such a characterization, in general, is difficult as there seems to be no direct way of obtaining a functional mapping $\overline{A}$ to $\mu^{\overline{A}}$. In fact, a much simpler situation (Kalman filtering with intermittent observations) involving a single sensor with observation packet losses demands the machinery of moderate deviations ([20]) and large random matrix theory ([22]) for a characterization of the invariant measures.

APPENDIX A
RANDOM DYNAMICAL SYSTEMS: FACTS AND RESULTS

We start by defining a random dynamical system (RDS). In the sequel, we follow the notation in [30], [27].

Definition 17 (RDS) A RDS with (one-sided) time $\mathbb{T}_+$ and state space $\mathcal{X}$ is a pair (θ, φ) with the following properties:

A) A metric dynamical system $\theta = (\Omega, \mathcal{F}, \mathbb{P}, \{\theta_t, t \in \mathbb{T}\})$ with two-sided time $\mathbb{T}$, i.e., a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a family of transformations $\{\theta_t : \Omega \longmapsto \Omega\}_{t \in \mathbb{T}}$ such that

A.1) $\theta_0 = \mathrm{id}_\Omega, \quad \theta_t \circ \theta_s = \theta_{t+s}, \quad \forall t, s \in \mathbb{T}$

A.2) $(t, \omega) \longmapsto \theta_t \omega$ is measurable.

A.3) $\theta_t \mathbb{P} = \mathbb{P} \quad \forall t \in \mathbb{T}$, i.e., $\mathbb{P}(\theta_t B) = \mathbb{P}$ for all $B \in \mathcal{F}$ and all $t \in \mathbb{T}$.

B) A cocycle φ over θ of continuous mappings of $\mathcal{X}$ with time $\mathbb{T}_+$, i.e., a measurable mapping

$$\varphi : \mathbb{T}_+ \times \Omega \times \mathcal{X}, \quad (t, \omega, X) \longmapsto \varphi(t, \omega, X) \tag{98}$$

such that

B.1) The mapping $X \longmapsto \varphi(t, \omega, X) \equiv \varphi(t, \omega)X$ is continuous in X for every $t \in \mathbb{T}_+$ and $\omega \in \Omega$.

B.2) The mappings $\varphi(t, \omega) \doteq \varphi(t, \omega, \cdot)$ satisfy the cocycle property:

$$\varphi(0, \omega) = \mathrm{id}_\mathcal{X}, \quad \varphi(t + s, \omega) = \varphi(t, \theta_s \omega) \circ \varphi(s, \omega) \tag{99}$$

for all $t, s \in \mathbb{T}_+$ and $\omega \in \Omega$.

Although we consider in this paper discrete time RDS, the general notion of RDS, as defined in [27], applies equally well to dynamical systems with continuous time. In the above definition, the randomness is captured by the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and iterates indexed by ω indicates pathwise construction. For example, if X_0 is the deterministic initial state of the system of interest at time $t = 0$, the random state at time $t \in \mathbb{T}_+$ is given by

$$X_t(\omega) = \varphi(t, \omega, X_0) \tag{100}$$

The measurability assumptions in the definition above, guarantee that the random state X_t is a well-defined random variable. Also, note that the iterates are defined for non-negative (one-sided) time, however, the family of transformations $\{\theta_t\}$ is two-sided, which is purely for technical convenience, as will be seen later.

Some results from RDS theory We summarize terminology and notions used in the RDS literature (see [26], [27] for details.)

Consider a generic RDS (θ, φ) with state space $\mathcal{X}$ as in Definition 17. In the following we assume that $\mathcal{X}$ is a non-empty subset of a real Banach space V with a closed, convex, solid, normal (w.r.t. the Banach space norm,) minihedral cone V_+. We denote by $\preceq$ the partial order induced by V_+ in $\mathcal{X}$ and $<<$ denotes the corresponding strong order. Although the development that follows may hold for arbitrary $\mathcal{X} \subset V$, in the sequel we assume $\mathcal{X} = V_+$ (which is true for the RDS (θ^R, φ^R) modeling the RARE.)

Definition 18 (Order-Preserving RDS) An RDS (θ, φ) with state space V_+ is called order-preserving if

$$X \preceq Y \implies \varphi(t, \omega, X) \preceq \varphi(t, \omega, Y), \quad \forall t \in \mathbb{T}_+, \omega \in \Omega, X, Y \in V_+ \tag{101}$$

Definition 19 (Sublinearity) An order-preserving RDS (θ, φ) with state space V_+ is called sublinear if for every $X \in V_+$ and $\lambda \in (0, 1)$ we have

$$\lambda \varphi(t, \omega, X) \preceq \varphi(t, \omega, \lambda X), \quad \forall t > 0, \ \omega \in \Omega \tag{102}$$

The RDS is said to be strictly sublinear if strict inequality in eqn. (102) holds for $X \in \mathrm{int}V_+$, i.e. for $X \in \mathrm{int}V_+$,

$$\lambda \varphi(t, \omega, X) \prec \varphi(t, \omega, \lambda X), \quad \forall t > 0, \ \omega \in \Omega \tag{103}$$

and strongly sublinear if in addition to eqn. (102), we have

$$\lambda \varphi(t, \omega, X) \ll \varphi(t, \omega, \lambda X), \quad \forall t > 0, \ \omega \in \Omega, \ X \in \mathrm{int}V_+ \tag{104}$$

Definition 20 (Equilibrium) A random variable $u : \Omega \longmapsto V_+$ is called an equilibrium (fixed point, stationary solution) of the RDS (θ, φ) if it is invariant under φ, i.e.,

$$\varphi(t, \omega, u(\omega)) = u(\theta_t\omega), \quad \forall t \in \mathbb{T}_+, \omega \in \Omega \tag{105}$$

In case, eqn. (105) holds for all $\omega \in \Omega$, except on a set of $\mathbb{P}$ measure zero, we call u an almost equilibrium.

Since, the transformations $\{\theta_t\}$ are measure-preserving, i.e., $\theta_t\mathbb{P} = \mathbb{P}$, $\forall t$, we have

$$u(\theta_t\omega) \overset{d}{=} u(\omega), \quad \forall t \tag{106}$$

Thus eqn. (105), in particular, implies that, for an almost equilibrium u, the sequence of iterates $\{\varphi(t, \omega, u(\omega))\}_{t \in \mathbb{T}_+}$ have the same distribution, which is the distribution of u.

Definition 21 (Part) The equivalence classes in V_+ under the equivalence relation defined by $X \sim Y$ if there exists $\alpha_0 \geq 1$ such that $\alpha_0^{-1}X \preceq Y \preceq \alpha_0 X$ are called parts of V_+.

We call the part C_v generated by a random variable $v : \Omega \longmapsto V_+$ as the collection of random variables $u : \Omega \longmapsto V_+$ such that there exists deterministic $\alpha_u \geq 1$ with

$$\alpha_u^{-1}v(\omega) \preceq u(\omega) \preceq \alpha_u v(\omega), \quad \forall \omega \in \Omega \tag{107}$$

Definition 22 (Orbit) For a random variable $u : \Omega \longmapsto V_+$ we define the *forward* orbit $\eta_u^f(\omega)$ emanating from $u(\omega)$ as the random set $\{\varphi(t, \omega, u(\omega))\}_{t \in \mathbb{T}_+}$. The forward orbit gives the sequence of iterates of the RDS starting at u.

Although η_u^f is the object of practical interest, for technical convenience (will be seen later,) we also define the *pull-back* orbit $\eta_u^b(\omega)$ emanating from u as the random set $\{\varphi(t, \theta_{-t}\omega, u(\theta_{-t}\omega))\}_{t \in \mathbb{T}_+}$.

The reason for defining the pull-back orbit is that it is comparatively convenient to establish asymptotic properties for η_u^b. However, analyzing η_u^b leads to understanding asymptotic distributional properties for η_u^f, because the random sequences $\{\varphi(t, \omega, u(\omega))\}_{t \in \mathbb{T}_+}$ and $\{\varphi(t, \theta_{-t}\omega, u(\theta_{-t}\omega))\}_{t \in \mathbb{T}_+}$ are equivalent in distribution. In other words,

$$\varphi(t, \omega, u(\omega)) \overset{d}{=} \varphi(t, \theta_{-t}\omega, u(\theta_{-t}\omega)), \quad \forall t \in \mathbb{T}_+ \tag{108}$$

This follows from the fact that $\theta_t\mathbb{P} = \mathbb{P}$, $\forall t \in \mathbb{T}$. Thus, in particular, we have the following assertion.

Lemma 23 Let the sequence $\{\varphi(t, \theta_{-t}\omega, u(\theta_{-t}\omega))\}_{t \in \mathbb{T}_+}$ converge in distribution to a measure μ on V_+, where $u : \Omega \longmapsto V_+$ is a random variable. Then the sequence $\{\varphi(t, \omega, u(\omega))\}_{t \in \mathbb{T}_+}$ also converges in distribution to the measure μ.

We now introduce notions of boundedness of RDS, which will be used in the sequel.

Definition 24 (Boundedness) Let $a : \Omega \longmapsto V_+$ be a random variable. The pull-back orbit $\eta_a^b(\omega)$ emanating from a is said to be bounded on $U \in \mathcal{F}$ is there exists a random variable C on U such that

$$\|\varphi(t, \theta_{-t}\omega, a(\theta_{-t}\omega))\| \leq C(\omega), \quad \forall t \in \mathbb{T}_+, \omega \in U \tag{109}$$

Definition 25 (Conditionally Compact RDS) An RDS (θ, φ) in V_+ is said to be conditionally compact if for any $U \in \mathcal{F}$ and pull-back orbit $\eta_a^b(\omega)$ which is bounded on U there exists a family of compact sets $\{K(\omega)\}_{\omega \in U}$ such that

$$\lim_{t \to \infty} \text{dist}(\varphi(t, \theta_{-t}\omega, a(\theta_{-t}\omega)), K(\omega)) = 0, \quad \omega \in U \tag{110}$$

It is to be noted that conditionally compact is a topological property of the space V_+.

We now state a limit set dichotomy result for a class of sublinear, order-preserving RDS.

Theorem 26 (Corollary 4.3.1. in [27]) Let V be a separable Banach space with a normal solid cone V_+. Assume that (θ, φ) is a strongly sublinear conditionally compact order-preserving RDS over an ergodic metric dynamical system θ. Suppose that $\varphi(t, \omega, 0) \gg 0$ for all $t > 0$ and $\omega \in \Omega$. Then precisely one of the following applies:

(a) For any $X \in V_+$ we have

$$\mathbb{P}\left(\lim_{t \to \infty} \|\varphi(t, \theta_{-t}\omega, X)\| = \infty\right) = 1 \tag{111}$$

(b) There exists a unique almost equilibrium $u(\omega) \gg 0$ defined on a θ-invariant set[12] $\Omega^* \in \mathcal{F}$ with $\mathbb{P}(\Omega^*) = 1$ such that for any random variable $v(\omega)$ possessing the property $0 \preceq v(\omega) \preceq \alpha_0 u(\omega)$ for all $\omega \in \Omega^*$ and deterministic $\alpha_0 > 0$, the following holds:

$$\lim_{t \to \infty} \varphi(t, \theta_{-t}\omega, v(\theta_{-t}\omega)) = u(\omega), \quad \omega \in \Omega^* \tag{112}$$

APPENDIX B
PROOFS IN SECTION V

Proof of Lemma 14 The proof is obtained by constructing an approximate filter with suboptimal performance, and then bounding its error by using the rank condition on the Grammian G_{w_0}. We detail such a construction now.

Consider ℓ steps $t = 1, \cdots, \ell$ of the linear time-varying signal/observation model given by 2 and 3, where in 3 we index the observation matrix as $\mathcal{C}_{n_t}$ where $n(t)$ indicates the current state of the walk w_0. The signal vector $\mathbf{x}_t \in \mathbb{R}^M$ with initial state $\mathbf{x}_1$ is a Gaussian random variable with known mean $\overline{\mathbf{x}}_1$ and variance $X \in \mathbb{S}_+^N$. The system noise process $\{\mathbf{w}_t\}$ is uncorrelated zero mean Gaussian with covariance $\mathcal{Q}$. The observation noise process $\{\mathbf{v}_t\}_{t=1}^\ell$ is uncorrelated zero mean Gaussian with time varying error covariance $\mathcal{R}_{n_t}$ and independent of the initial signal state and the system noise process. By the above construction, the optimal estimate of the signal state $\mathbf{x}_t$ at time t, based on observations till that time, is given by the Kalman filter initialized with X as the predicted conditional error covariance at time $t = 1$. In other words, the optimal m.m.s.e. state estimator (predictor form)

$$\widehat{\mathbf{x}}_{w_0}(t) = \mathbb{E}\left[\mathbf{x}_t \,|\, \{\mathbf{y}_s\}_{1 \leq s < t}\right] \tag{113}$$

of $\mathbf{x}_t$ based on observations $\{\mathbf{y}_s\}_{1 \leq s < t}$ for $1 \leq t \leq \ell + 1$ can be recursively constructed through the Kalman filter and the corresponding predicted conditional error covariance sequence $\{P_{w_0}(t)\}_{1 \leq t \leq \ell+1}$ satisfies the recursion:

$$P_{w_0}(t+1) = \mathcal{F}P_{w_0}(t)\mathcal{F}^T + \mathcal{Q} - \mathcal{F}P_{w_0}(t)\mathcal{C}_{n_t}^T\left(\mathcal{C}_{n_t}P_{w_0}(t)\mathcal{C}_{n_t}^T + \mathcal{R}_{n_t}\right)^{-1}\mathcal{C}_{n_t}P_{w_0}(t)\mathcal{F}^T \tag{114}$$

with initial condition $P_{w_0}(1) = X$. We then have

$$P_{w_0}(\ell+1) = f_{n_\ell} \circ \cdots \circ f_{n_1}(X) \tag{115}$$

the R.H.S. being the desired functional form in eqn. (40), i.e., $P_{w_0}(\ell+1) = g_{w_0}(X)$. Since for a Kalman filter with deterministic system/observation matrices, the conditional error covariance is equal to the unconditional one and the fact that the Kalman filter minimizes any positive definite form of the estimation error, for a generic estimator $\widehat{\mathbf{h}}$ of $\mathbf{x}_{\ell+1}$ based on $\{\mathbf{y}_s\}_{1 \leq s \leq \ell}$ we have

$$P_{w_0}(\ell+1) \preceq \mathbb{E}\left[\left(\mathbf{x}_{\ell+1} - \widehat{\mathbf{h}}\right)\left(\mathbf{x}_{\ell+1} - \widehat{\mathbf{h}}\right)^T\right] \tag{116}$$

where $\preceq$ refers to the partial order on $\mathbb{S}_+^N$. In order to upper bound the functional g_{w_0}, we now construct a suboptimal state estimator with a guaranteed estimation performance. To this end, define the modified Grammian

$$\widetilde{G}_{w_0} = \sum_{t=1}^\ell \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t}^T \mathcal{R}_{n_t}^{-1} \mathcal{C}_{n_t} \mathcal{F}^{t-1} \tag{117}$$

---

[12]A set $A \in \mathcal{F}$ is called θ-invariant if $\theta_t A = A$ for all $t \in \mathbb{T}$.

We note that $\widetilde{G}_{w_0}$ is invertible by the invertibility of G_{w_0} and the noise covariances $\mathcal{R}_{n_t}$. Define the suboptimal estimator of $\mathbf{x}_{\ell+1}$ by:

$$\overline{\mathbf{x}}_{w_0}(\ell+1) = \mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathbf{y}_t \tag{118}$$

based on observations $\{\mathbf{y}_s\}_{1 \le s \le \ell}$. Using the fact, that,

$$\mathbf{x}_t = \mathcal{F}^{t-1}\mathbf{x}_1 + \sum_{s=1}^{t-1} \mathcal{F}^{t-1-s}\mathbf{w}_s, \quad 1 \le t \le \ell+1 \tag{119}$$

we have from eqn. (118)

$$\begin{aligned}
\overline{\mathbf{x}}_{w_0}(\ell+1) &= \mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathcal{C}_{n_t} \mathcal{F}^{t-1} \mathbf{x}_1 + \mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathcal{C}_{n_t} \sum_{s=1}^{t-1} \mathcal{F}^{t-1-s}\mathbf{w}_s \\
&\quad + \mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathbf{v}_{n_t} \\
&= \mathcal{F}^\ell \mathbf{x}_1 + \mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathcal{C}_{n_t} \sum_{s=1}^{t-1} \mathcal{F}^{t-1-s}\mathbf{w}_s + \mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathbf{v}_{n_t}
\end{aligned}$$

The filtering error is then given by

$$\begin{aligned}
\mathbf{e}_{w_0}(\ell+1) &= \mathbf{x}_{\ell+1} - \overline{\mathbf{x}}_{w_0}(\ell+1) \\
&= -\mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathcal{C}_{n_t} \sum_{s=1}^{t-1} \mathcal{F}^{t-1-s}\mathbf{w}_s - \mathcal{F}^\ell \widetilde{G}_{w_0}^{-1} \sum_{t=1}^{\ell} \left(\mathcal{F}^{t-1}\right)^T \mathcal{C}_{n_t} \mathcal{R}_{n_t}^{-1} \mathbf{v}_{n_t} \tag{120}
\end{aligned}$$

We note that the error above is independent of the initial state $\mathbf{x}_1$ (and hence the covariance X) of the system, and the mean square boundedness of the process noise $\{\mathbf{w}_t\}$ and observation noise $\{\mathbf{v}_t\}$ imply the existence of a constant $\alpha_0 > 0$, such that,

$$\mathbb{E}\left[\mathbf{e}_{w_0}(\ell+1)\mathbf{e}_{w_0}(\ell+1)^T\right] \preceq \alpha_0 I \tag{121}$$

The Lemma then follows by the optimality of the Kalman filter, as stated in eqn. (116).

Proof of Lemma 15 In case $\mathcal{F}$ is stable, the claim is obvious, as the suboptimal estimate of 0 at each sensor for all time is stochastically bounded. So, in the sequel we assume $\mathcal{F}$ is unstable.

The proof is somewhat technical and mainly uses the uniform boundedness of the composition of Riccati operators in Lemma 14 and the ergodicity of the underlying switching Markov chain $\{\widetilde{p}(t)\}_{t \in \mathbb{T}_+}$. From Lemma 14 it follows that a successive application of ℓ Riccati maps (in the composition order $f_{n_l} \circ \cdots \circ f_{n_1}$) reduces the iterate in the conic interval $[0, \alpha_0 I]$ irrespective of its initial value. The approach is to relate the probability of large exceedance of $\widetilde{P}_t$ to the hitting time statistics of a modified Markov chain. We detail it below.

First, we note that the regularity of the distributions of $\widetilde{P}(t)$ for every t, implies that it suffices to show

$$\lim_{J \to \infty} \sup_{t \ge t_0} \mathbb{P}\left(\left\|\widetilde{P}(t)\right\| > J\right) = 0 \tag{122}$$

for some arbitrarily large $t_0 \in \mathbb{T}_+$. For every n, the Riccati update is upper bounded by the Lyapunov operator, i.e.,

$$f_n(X) \preceq \mathcal{F}X\mathcal{F}^T + \mathcal{Q}, \quad \forall X \in \mathbb{S}_+^N \tag{123}$$

For sufficiently large $J > 0$, define

$$k(J) = \max_k \left\{ k \in \mathbb{T}_+ \,\middle|\, \alpha^{2k}\alpha_0 + \frac{\alpha^{2k}-1}{\alpha^2-1}\|\mathcal{Q}\| \le J \right\} \tag{124}$$

where $\alpha = \|\mathcal{F}\|$. Since $\mathcal{F}$ is unstable ($\alpha > 1$), we note that $k(J) \to \infty$ as $J \to \infty$.

We introduce additional notation here. For integers $t_0, t_1 \geq \ell$, the phrase "there exists a $(n_1, n_2, \cdots, n_\ell)$ cycle in the interval $[t_0, t_1]$" indicates the existence of an integer $t_0 \leq \acute{t} \leq t_1$, such that,

$$\widetilde{p}(\acute{t} - \ell + s) = n_s, \quad 1 \leq s \leq \ell \tag{125}$$

where $\{\widetilde{p}(t)\}_{t \in \mathbb{T}_+}$ is the switching Markov chain.

We now make the following claim for relating the probabilities of interest for sufficiently large t:

$$\mathbb{P}\left(\left\|\widetilde{P}(t)\right\| > J\right) \leq \mathbb{P}\left(\text{no } (n_1, n_2, \cdots, n_\ell) \text{ exists in } [t - k(J), t]\right) \tag{126}$$

Indeed, assume on the contrary that a $(n_1, \cdots, n_\ell)$ cycle exists in the interval $[t - k(J), t]$. Then there exists $\acute{t} \in [t - k(J), t]$, such that,

$$\widetilde{p}(\acute{t} - \ell + s) = n_s, \quad 1 \leq s \leq \ell \tag{127}$$

This implies

$$\widetilde{P}\left(\acute{t}\right) = f_{n_\ell} \circ \cdots \circ f_{n_1}\left(\widetilde{P}\left(\acute{t} - \ell + 1\right)\right) \tag{128}$$

and hence by Lemma 14

$$\widetilde{P}\left(\acute{t}\right) \preceq \alpha_0 I \tag{129}$$

which holds irrespective of the value of $\widetilde{P}\left(\acute{t} - \ell + 1\right)$. By eqn. (123) we note that

$$\widetilde{P}(s) \preceq \mathcal{F}\widetilde{P}(s - 1)\mathcal{F}^T + \mathcal{Q}, \quad \forall s \tag{130}$$

Continuing the recursion and noting $\widetilde{P}\left(\acute{t}\right) \preceq \alpha_0 I$

$$\left\|\widetilde{P}(t)\right\| \leq \alpha^{2\left(t - \acute{t}\right)}\left\|\widetilde{P}\right\| + \frac{\alpha^{2\left(t - \acute{t}\right)} - 1}{\alpha^2 - 1}\|\mathcal{Q}\| = \alpha^{2\left(t - \acute{t}\right)}\alpha_0 + \frac{\alpha^{2\left(t - \acute{t}\right)} - 1}{\alpha^2 - 1}\|\mathcal{Q}\|$$

Since $\left(t - \acute{t}\right) \leq k(J)$, it follows from the above

$$\left\|\widetilde{P}(t)\right\| \leq \alpha^{2\left(t - \acute{t}\right)}\alpha_0 + \frac{\alpha^{2\left(t - \acute{t}\right)} - 1}{\alpha^2 - 1}\|\mathcal{Q}\| \leq \alpha^{2k(J)}\alpha_0 + \frac{\alpha^{2k(J)} - 1}{\alpha^2 - 1}\|\mathcal{Q}\| \leq J$$

where the last step follows from the definition of $k(J)$ (eqn. (124)). We thus note that the existence of a $(n_1, \cdots, n_\ell)$ cycle in $[t - k(J), t]$ implies $\left\|\widetilde{P}(t)\right\| \leq J$. i.e., we have the event inclusion:

$$\{\text{there exists a } (n_1, \cdots, n_\ell) \text{ cycle in } [t - k(J), t]\} \subset \left\{\left\|\widetilde{P}(t)\right\| \leq J\right\} \tag{131}$$

The claim in eqn. (126) follows. Thus estimating the probability on the L.H.S. of eqn. (126) reduces to estimating the probability of a $(n_1, \cdots, n_\ell)$ cycle in $[t - k(J), t]$. To this end we construct another Markov chain $\{z(t)\}_{t \geq \ell}$. The state space $\mathcal{Z}$ is a subset of V^ℓ given by:

$$\mathcal{Z} = \left\{z = (i_1, i_2, \cdots, i_\ell) \mid \overline{A}_{i_j, i_{j+1}} > 0, \quad 1 \leq j < \ell\right\} \tag{132}$$

The dynamics of the Markov chain $\{z(t)\}_{t \geq \ell}$ is given in terms of the Markov chain $\{\widetilde{p}(t)\}_{t \in \mathbb{T}_+}$ as follows:

$$z(t) = (\widetilde{p}(t - \ell + 1), \widetilde{p}(t - \ell + 2), \cdots, \widetilde{p}(t)) \tag{133}$$

From the dynamics of $\{\widetilde{p}(t)\}_{t \in \mathbb{T}_+}$ it follows that $\{z(t)\}_{t \geq \ell}$ is a Markov chain with transition probability $\overline{A}_{nl}$ between allowable states $(i_1, i_2, \cdots, i_{\ell-1}, n)$ and $(i_2, \cdots, 1_\ell, n, l)$. With state space $\mathcal{Z}$, the Markov chain $\{z(t)\}$ inherits irreducibility and aperiodicity from that of $\{\widetilde{p}(t)\}$. Also, $\{z(t)\}$ is stationary from the stationarity of $\{\widetilde{p}(t)\}$

with invariant distribution:

$$\mathbb{P}\left(z(t) = (i_1, i_2, \cdots, i_\ell)\right) = \frac{1}{N} \prod_{j=1}^{\ell-1} \overline{A}_{j,j+1}, \quad (i_1, i_2, \cdots, i_\ell) \in \mathcal{Z}, \ t \geq \ell, \ t \in \mathbb{T}_+ \tag{134}$$

Denote the hitting time τ_0 of $\{z(t)\}$ to the state $(n_1, \cdots, n_\ell)$ by:

$$\tau_0 = \min\{t > \ell \, | \, z(t) = (n_1, \cdots, n_\ell)\} \tag{135}$$

and for all $z \in \mathcal{Z}$ define

$$\mathbb{P}_z\left(\tau_0 > s\right) = \mathbb{P}\left(\tau_0 > s \mid z(\ell) = z\right) \tag{136}$$

Also, for each $t \geq \ell$ and J sufficiently large, define the stopping times

$$\tau_t^J = \min\{t \geq t - k(J) \, | \, z(t) = (n_1, \cdots, n_\ell)\} \tag{137}$$

From the Markov property it then follows

$$\mathbb{P}\left(\tau_t^J > t \mid z(t - k(J) - 1) = z\right) = \mathbb{P}_z\left(\tau_0 > k(J) + 1\right) \tag{138}$$

It then follows successively

$$
\begin{aligned}
\mathbb{P}\left(\text{no } (n_1, \cdots, n_\ell) \text{ exists in } [t - k(J), t]\right) &= \mathbb{P}\left(\tau_t^J > t\right) \\
&= \sum_{z \in \mathcal{Z}} [\mathbb{P}\left(z(t - k(J) - 1) = z\right) \\
&\quad \mathbb{P}\left(\tau_t^J > t \mid z(t - k(J) - 1) = z\right)] \\
&= \sum_{z \in \mathcal{Z}} \mathbb{P}\left(z(t - k(J) - 1) = z\right) \mathbb{P}_z\left(\tau_0 > k(J) + 1\right)
\end{aligned}
\tag{139}
$$

Since the above development holds for all $t \geq t_0$ for some sufficiently large t_0, we conclude from eqn. (126)

$$\sup_{t \geq t_0} \mathbb{P}\left(\left\|\widetilde{P}(t)\right\| > J\right) \leq \sum_{z \in \mathcal{Z}} \mathbb{P}\left(z(t - k(J) - 1) = z\right) \mathbb{P}_z\left(\tau_0 > k(J) + 1\right) \tag{140}$$

The recurrence (in fact positive recurrence) of the finite state Markov chain $\{z(t)\}$ and the fact that $k(J) \to \infty$ as $J \to \infty$ imply, for all $z \in \mathcal{Z}$,

$$\lim_{J \to \infty} \mathbb{P}_z\left(\tau_0 > k(J) + 1\right) = 0 \tag{141}$$

Since $\mathcal{Z}$ is finite, letting $J \to \infty$ in eqn. (140) leads to

$$\lim_{J \to \infty} \sup_{t \geq t_0} \mathbb{P}\left(\left\|\widetilde{P}(t)\right\| > J\right) = 0 \tag{142}$$

by the dominated convergence theorem and the Lemma follows.

REFERENCES

[1] B. Rao and H. Durrant-Whyte, "Fully decentralized algorithm for multisensor Kalman filtering," in *Inst. Electr. Eng.-Control Theory Appl.*, vol. 138, September 1991, pp. 413–420.

[2] R. Olfati-Saber, "Distributed Kalman filter with embedded consensus filters," in *CDC-ECC'05, 44th IEEE Conference on Decision and Control and 2005 European Control Conference*, 2005, pp. 8179–8184.

[3] U. A. Khan and J. M. F. Moura, "Distributing the Kalman filter for large-scale systems," *IEEE Transactions on Signal Processing*, vol. 56, no. 10, pp. 4919–4935, October 2008.

[4] R. Carli, A. Chiuso, L. Schenato, and S. Zampieri, "Distributed Kalman filtering using consensus strategies," *IEEE Journal on Selected Areas of Communications*, vol. 26, no. 4, pp. 622–633, September 2008.

[5] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, pp. 2508 – 2530, June 2006.

[6] N. McKeown, A. Mekkittikul, V. Anantharam, and J. Walrand, "Achieving 100 % throughput in an input-queued switch," *IEEE Transactions on Communications*, vol. 47, no. 8, pp. 1260–1267, August 1999.

[7] G. K. Batchelor, *An introduction to fluid dynamics*, ser. Cambridge Mathematical Library. Cambridge University Press, 1973.

[8] H. Chen, P. Kumar, and J. van Schuppen, "On Kalman filtering for conditionally Gaussian systems with random matrices," *Systems & Control Letters*, vol. 13, pp. 397–404, 1989.

[9] Y. Wang and L. Guo, "On stability of random Riccati equations," *Science in China Series E: Technological Sciences*, vol. 42, no. 2, pp. 136–148, April 1999.

[10] A. Matveev and A. Savkin, "The problem of state estimation via asynchronous communication channels with irregular transmission times," *IEEE Transactions on Automatic Control*, vol. 48, pp. 670–676, 2006.

[11] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. Jordan, and S. Sastry, "Kalman filtering with intermittent observations," *IEEE Transactions of Automatic Control*, vol. 49, no. 9, pp. 1453–1464, 2004.

[12] X. Liu and A. Goldsmith, "Kalman filtering with partial observation losses," in *Proceedings of IEEE Conference on Decision and Control*, vol. 4, Bahamas, December 2004, pp. 4180–4186.

[13] V. Gupta, T. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, February 2006.

[14] Y. Xu and J. Hespanha, "Estimation under controlled and uncontrolled communications in networked control systems," in *IEEE Conference on Decision and Control*, Sevilla, Spain, December 2005, pp. 842–847.

[15] M. Huang and S. Dey, "Stability of Kalman filtering with Markovian packet losses," *Automatica*, vol. 43, no. 4, pp. 598–607, 2007.

[16] K. Plarre and F. Bullo, "On Kalman filtering for detectable systems with intermittent observations," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 386–390, February 2009.

[17] C. Robinson and P. R. Kumar, "Sending the most recent observation is not optimal in networked control: Linear temporal coding and towards the design of a control specific transport protocol," in *46th IEEE Decision and Control Conference*, New Orleans, LA, Dec 2007, pp. 334–339.

[18] L. Xie and L. Xie, "Stability of a random Riccati equation with Markovian binary switching," in *46th IEEE Decision and Control Conference*, New Orleans, LA, Dec 2007, pp. 1565–1570.

[19] S. Kar, B. Sinopoli, and J. M. F. Moura, "Kalman filtering with intermittent observations: Weak convergence to a stationary distribution," August 2009, accepted for publication in the IEEE Transactions on Automatic Control with mandatory revisions. [Online]. Available: http://arxiv.org/abs/0903.2890

[20] S. Kar and J. M. F. Moura, "Kalman filtering with intermittent observations: Weak convergence and moderate deviations," October 2009, submitted for publication. [Online]. Available: http://arxiv.org/abs/0910.4686

[21] A. Censi, "On the performance of Kalman filtering with intermittent observations: a geometric approach with fractals," in *American Control Conference (ACC)*, 2009, pp. 3806–3812.

[22] A. Vakili and B. Hassibi, "A Stieltjes transform approach for studying the steady-state behavior of random Lyapunov and Riccati recursions," in *47th IEEE Conference on Decision and Control, CDC 2008*, Cancún, Mexico, Dec 2008, pp. 453–458.

[23] P. Bougerol, "Kalman filtering with random coefficients," *SIAM Journal on Control and Optimization*, vol. 31, no. 4, pp. 942 – 959, July 1993.

[24] J. Jacod and A. Shiryaev, *Limit theorems for stochastic processes*. Berlin Heidelberg: Springer-Verlag, 1987.

[25] P. Diaconis and D. Freedman, "Iterated random functions," *SIAM REVIEW*, vol. 41, no. 1, pp. 45–76, 1999.

[26] L. Arnold, *Random Dynamical Systems*. Springer, 1998.

[27] I. Chueshov, *Monotone Random Systems - Theory and Applications*. Springer, 2002.

[28] O. Kallenberg, *Foundations of Modern Probability*, 2nd ed. Springer Series in Statistics., 2002.

[29] S. Ethier and T. Kurtz, *Markov Processes: Characterization and Convergence*. John Wiley and Sons, Inc., 1986.

[30] L. Arnold and I. Chueshov, "Order-preserving random dynamical systems: Equilibria, attractors, applications," *Dynamics and Stability of Systems*, vol. 13, pp. 265–280, 1998.